

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Massenspektrometrische Evaluierung des DNA und RNA Metabolismus

Jessica Anna Steinbacher

aus

Freising

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München den 06. November 2016

Jessica Anna Steinbacher

Dissertation eingereicht am:

13.12.2016

1. Gutachter:

Prof. Dr. T. Carell

2. Gutachter:

PD Dr. Stylianos Michalakis

Mündliche Prüfung am:

20.01.2017

Danksagung

An dieser Stelle möchte ich mich bei zahlreichen Personen für ihre Unterstützung während meiner Promotion bedanken.

Mein erster Dank geht an meinen Doktorvater *Prof. Dr. Thomas Carell* für die spannende Themenauswahl, die Unterstützung und die hervorragenden Arbeitsbedingungen.

Herrn *PD Dr. Stylianos Michalakis* danke ich für die Übernahme des Koreferats.

Auch möchte ich mich den Mitgliedern der Prüfungskommission herzlich bedanken.

Frau *Slava Gärtner* danke ich für ihre Hilfe bei allen organisatorischen Belangen. Auch möchte ich mich ganz herzlich bei Frau *Kerstin Kurz* bedanken, die mir stets eine sehr große Hilfe bei der Bewältigung riesiger Probenmengen für die massenspektrometrische Analyse war.

Markus Müller danke ich für seine Hilfsbereitschaft, für seine Geduld beim Erklären neuer biochemischer Methoden, für spannende Diskussionen und für die Korrektur dieser Arbeit.

Mein besonderer Dank geht an alle mit denen ich an verschiedensten Projekten zusammen arbeiten durfte. Ich danke *Olesea Kosmatchev* und *Katharina Iwan* für die großartige Zusammenarbeit am Triple Quad und bei der Erforschung verschiedenster aktiver DNA-Demethylierungswege.

Ein besonderer Dank geht auch Simon Geiger, der nie aufgehört hat an HNRNPK zu glauben, sowie den Rest der RNA Bande, *Martin Rossa*, *Matthias Kurz*, *Katharina Iwan*, *Olga Kotljarowa*, *Angie Kirchner* und *Sarah Schiffers*.

Zudem danke ich *Arne Schröder*, *Rene Rahimoff*, *Fabio Spada*, *Mirko Wagner*, *Toni Pfaffeneder*, *Silvia Laube*, *David Eisen*, *Michael Stadlmeier* und *Andrea Künzel* für ihre Hilfe und die gute Zusammenarbeit. Arshan Perera und Bernard Thienpont möchte ich zudem für sehr interessante und fruchtbare gemeinsame Kooperationsprojekte danken.

Ein großer Dank geht auch an alle Korrekturleser, *Katharina Iwan*, *Olesea Kosmatchev*, *Fabio Spada*, *Rene Rahimoff*, *Benjamin Hackner*, *Angie Kirchner*, *Simon Geiger* und *Markus Müller*.

Auch meinem Labor möchte ich danken. Die gute Stimmung, exzellente Musikauswahl und angeregte Diskussionen haben den Laboraltag viel besser gemacht.

An dieser Stelle möchte ich den Betreuern meiner Masterarbeit, *Benjamin Hackner*, *Toni Pfaffeneder* und *Martin Münzel* danken, die mich sozusagen in die Gruppe und zum Thema Epigenetik geholt haben, sowie meinen Masterstudenten *Anna Casas*, *Katharina Iwan* und *Matthias Kurz*, ich habe viel von und mit euch gelernt.

Viele Kollegen aus der Carell Gruppe sind mittlerweile Freunde, ich bin sehr froh euch hier kennengelernt zu haben. Ich blicke auf unzählige lustige Abende beim Kochen, Topmodel gucken, im Biergarten, beim Grillen und auf Spieleabende zurück.

Auch meinen Freunden außerhalb der Uni möchte ich für ihre Unterstützung, Geduld und ihr Verständnis bei langen Arbeitstagen danken.

Mein größter Dank gilt meinen Eltern Chris und Flo, meinem Bruder Max und meinem Freund Ben, ohne eure Unterstützung, Liebe und gute Ratschläge hätte ich diese Zeit nicht so gut gemeistert, ich bin unendlich dankbar euch in meinen Leben zu haben.

Veröffentlichungen:

- [1] B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S.K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell. Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction, *Org. Lett.* **2013**, *15*, 366-369.
- [2] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A.S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K.T. Wanner, C. Ochsenfeld, T. Carell. Deamination, Oxidation and C-C bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxycytosine, *J Am Chem Soc.* **2013**, *135*, 14593-9.
- [3] A.S. Schröder, J. Steinbacher, B. Steigenberger, F.A. Gnerlich, S. Schiesser, T. Pfaffeneder, T. Carell. Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-Deoxycytidine, *Angew. Chem. Int. Ed.* **2014**, *53*, 315-8.
- [4] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller & T. Carell. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA, *Nat. Chem. Biol.* **2014**, *10*, 574-581.
- [5] A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis. TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression, *Cell Rep.* **2015**, *11*, 1-12.
- [6] M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell. , *Angew. Chem. Int. Ed.* **2015**, *54* (42), 12511-12514.
- [7] N. C. Gassen, G. R. Fries , A. S. Zannas , J. Hartmann , J. Zschocke , K. Hafner , T. Carrillo-Roa , J. Steinbacher , S. N. Preißinger , L. Hoeijmakers , M. Knop , F. Weber , S. Kloiber , S. Lucae , G.P. Chrousos , T. Carell , M. Ising , E. B. Binder , M. V. Schmidt , J. Rüegg , T. Rein. , *Sci. Signal.* **2015**, *8* (404).
- [8] B. Thienpont, J. Steinbacher, H. Zhao, F. D'Anna, A. Kuchnio, A. Ploumaki, B. Ghesquière, L. Van Dyck, B. Boeckx, L. Schoonjans, E. Hermans, F. Amant, V. N. Kristensen, K. P. Koh, M. Mazzone, M. L. Coleman, T. Carell, P. Carmeliet, D. Lambrechts. Tumour hypoxia causes DNA hypermethylation by reducing TET activity., *Nature*, **2016**, *537*, 63-68.

Inhaltsverzeichnis

Zusammenfassung.....	1
1 Einleitung.....	6
1.1 Modifikationen von DNA	6
1.1.1 Epigenetik.....	6
1.1.2 Ethenoaddukte in DNA.....	20
1.2 RNA Modifikationen	25
1.2.1 Bildung und Vorkommen der Isopentenylmodifikation ms ² i ⁶ A.....	27
1.2.2 Funktion von ms ² i ⁶ A.....	29
1.2.3 Rolle der Modifikation ms ² i ⁶ A in Krankheiten.....	31
2 Aufgabenstellung.....	34
3 Veröffentlichte Arbeiten	36
3.1 Altersabhängige Level von 5-Methyl-, 5-Hydroxymethyl- und 5-Formylcytosin in Hirngeweben des Menschen und der Maus	36
3.2 TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression	42
3.3 Tet enzymes oxidize T to hmU in mouse embryonic stem cells.....	57
3.4 Synthese eines DNA-Promotorsegments mit allen vier epigenetischen Nukleosiden: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl- und 5-Carboxy-2'-Desoxycytidin.....	69
3.5 Chaperoning epigenetics: FKBP51 decreases the activity of DNMT1 and mediates epigenetic effects of the antidepressant paroxetine	74
3.6 Tumor hypoxia causes DNA hypermethylation by reducing TET activity.....	89
4 Unveröffentlichte Arbeiten.....	116
4.1 Quantifizierung von Ethenoaddukten.....	116
4.1.1 Prolog.....	116
4.1.2 Ergebnisse und Diskussion.....	117
4.1.4 Zusammenfassung und Projektbeiträge	123
4.1.5 Material und Methoden.....	124
4.2 Decarboxylierungsstudien mit D ₂ - ¹⁵ N ₂ -dC	128
4.2.1 Prolog.....	128

4.2.2 Ergebnisse und Diskussion.....	130
4.2.3 Zusammenfassung und Projektbeiträge	144
4.2.4 Material und Methoden.....	145
4.3 Desaminierung von Methylcytosin zu Thymin	147
4.3.1 Prolog.....	147
4.3.2 Ergebnisse und Diskussion.....	149
4.3.4 Zusammenfassung und Projektbeiträge	165
4.3.5 Material und Methoden.....	166
4.4 Interaktoren und Biologische Funktion von ms ² i ⁶ A.....	170
4.4.1 Prolog.....	170
4.4.2 Ergebnisse und Diskussion.....	172
4.4.4 Zusammenfassung und Projektbeiträge	188
4.4.5 Material und Methoden.....	189
5 Abkürzungsverzeichnis.....	201
6 Literaturverzeichnis	206

Zusammenfassung

Im Jahr 2009 wurde die epigenetische Modifikation 5-Hydroxymethyl-2'-desoxycytidin (5hmdC) im Genom von Säugetieren entdeckt und als sechste DNA-Base deklariert. Zwei Jahre später konnten die höher oxidierten Spezies 5-Formyl-2'-desoxycytidin (5fdC) und 5-Carboxy-2'-desoxycytidin (5cadC) entdeckt werden. Während man von 5hmdC inzwischen weiß, dass es eine wichtige Rolle in der Kontrolle der Genexpression spielt, ist die Funktion der höher oxidierten Spezies immer noch unklar.

Zu Beginn der Doktorarbeit wurden zu diesem Zweck Organe der Maus hinsichtlich der Mengen an allen epigenetischen Modifikationen untersucht und mit dem Modifikationsgehalt von embryonalen Stammzellen der Maus (mESC) verglichen. (Abb.1A+B) Im Rahmen dieser Studie wurden auch die Level der Modifikationen 5hmdC (C) und 5fdC (D) während des Alterungsprozesses von Mäusen bestimmt. (Abb.1 C+D)

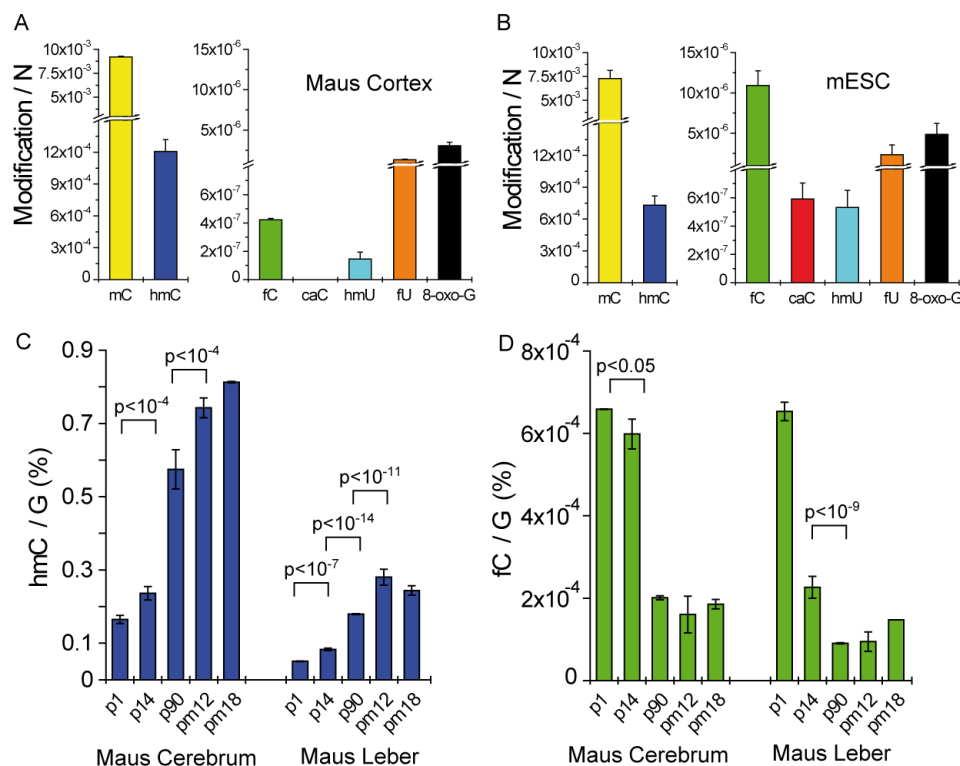


Abbildung 1: Absolute Quantifizierung der epigenetischen Modifikationen im Cortex der Maus (A) und in embryonalen Stammzellen der Maus (B). Bestimmung der Level an 5hmdC (C) und 5fdC (D) während des Alterungsprozesses im Cerebrum und der Leber von Mäusen.

Hier konnte festgestellt werden, dass die 5fdC Levels im Gegensatz zu 5hmdC während des Alterungsprozesses abnehmen. Somit scheint 5fdC eher zu Beginn des Lebens eine Rolle zu spielen, was auf eine Funktion bei der aktiven DNA-Demethylierung hindeuten könnte.

Im weiteren Verlauf der Arbeit wurde eine bereits bestehende Methode zur sensitiven Quantifizierung der epigenetischen Modifikationen so verändert, dass auch sehr geringe Mengen in verschiedenen Zelllinien und Organpräparationen quantifiziert werden konnten. Die Methode diente, durch die Quantifizierung von 5hmdC, zum einen der Bestätigung, dass der als spezifischer Tet 3 Binder identifizierte Transkriptionsrepressor REST dazu in der Lage ist, die Oxdaseaktivität von Tet 3 und damit die Levels an 5hmdC zu erhöhen. Zum anderen konnten ein Zusammenhang zwischen dem molekularen Chaperon Bindeprotein FKBP51, welches mit psychischen Erkrankungen in Verbindung gebracht wird, und der DNA Methyltransferase Dnmt1 hergestellt werden. So scheint FKBP51 über CDK5 eine Aktivierung von Dnmt1 zu bewirken und damit einen direkten Einfluss auf die Expression bestimmter Gene zu haben. Dies konnte über die Quantifizierung der 5mdC-Level bestätigt werden.

Im weiteren Verlauf der Arbeit konnte 5-Hydroxymethyl-2'-desoxyuridin (5hmdU) im Genom von Säugetieren in repräsentativen Mengen detektiert (siehe Abb. 1A+B) und der Ursprung der Modifikation aufgeklärt werden. So entsteht 5hmdU nur zu sehr geringen Anteilen durch eine Desaminierung von 5hmdC und vor allem aber durch eine Oxidation von dT durch die Tet Enzyme.

Die sensitive Quantifizierungsmethode konnte ebenfalls genutzt werden, um sehr geringe Mengen an 5mdC und 5hmdC in verschiedenen Tumorzelllinien zu bestimmen. Dadurch konnte gezeigt werden, dass die Mangelversorgung an Sauerstoff in Tumorzellen die Aktivität der Sauerstoff abhängigen Tet Enzyme reduziert. Dies tritt unabhängig von Veränderungen in der Tet Genexpression, dem Basalmetabolismus, der HIF Aktivität oder nukleären reaktiven Sauerstoffspezies auf und ist lediglich von der Knappheit an Sauerstoff beeinflusst. Die durch Hypoxie reduzierte Aktivität der Tet Enzyme verstärkt die Hypermethylierung an Promotoren. Die betroffenen Gene sind häufig an der DNA Reparatur, der Regulation des Zellzyklus, an der Angiogenese und an der Metastasierung beteiligt. Somit scheint die Hypoxie in Tumoren als neuartiger Regulator zu fungieren, welcher der DNA Methylierung unterliegt.

Wie schon erwähnt könnten die oxidierten Spezies 5fdC und 5cadC als Intermediate der

aktiven DNA Demethylierung fungieren. Zu diesem Zweck wurde eine Quantifizierungsmethode entwickelt, durch welche man mit Hilfe von Isotopenverfolgung Basenexzisionsreparatur vermittelte Demethylierung von direkter Demethylierung durch Deformylierung oder Decarboxylierung unterscheiden kann. Durch Fütterung von D_2 - $^{15}N_2$ -dC in mESCs konnte mit Hilfe von Isotopenverfolgung durch einen Vergleich von Tdg +/- Zellen mit Tdg -/- Zellen, in welchen nur noch der Weg der direkten aktiven Demethylierung möglich ist (siehe Abb. 2) erste Hinweise auf aktive Demethylierung über Deformylierung oder Decarboxylierung erhalten werden.

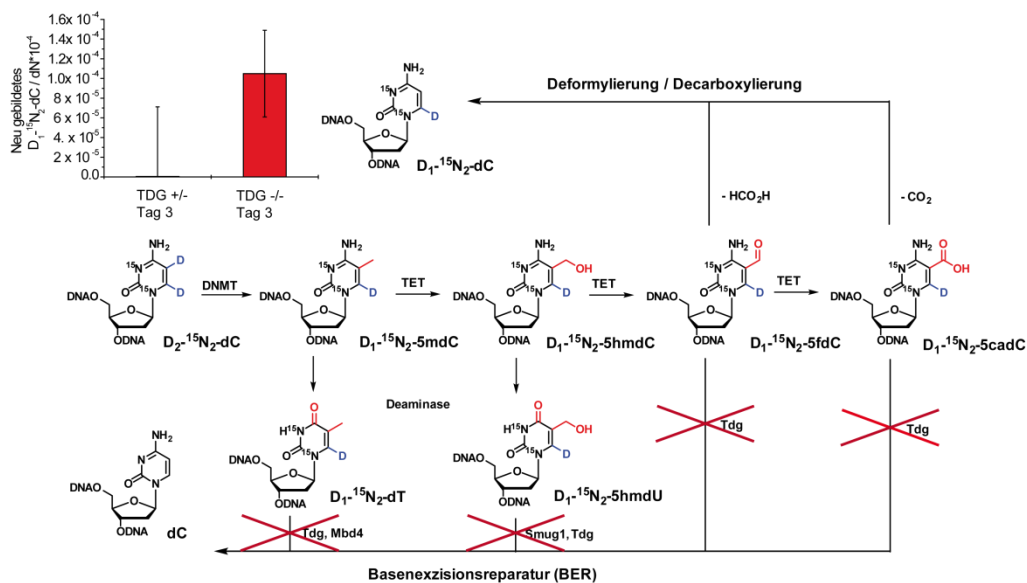


Abbildung 2: Mögliche Wege der oxidativen aktiven DNA-Demethylierung. Durch einen Knockout von Tdg kann vermehrt aktive Demethylierung durch Deformylierung und Decarboxylierung beobachtet werden.

In Tdg -/- Zellen war nach 3 Tagen eine Zunahme der Spezies D_1 - $^{15}N_2$ -dC zu verzeichnen, welche möglicherweise durch Deformylierung oder Decarboxylierung entstanden ist. Eine weitere Möglichkeit der aktiven Demethylierung von DNA ergibt sich aus der Desaminierung von 5mdC zu dT mit anschließender Basenexzisionsreparatur. Daher wurde eine Methode zur sensitiven Quantifizierung von isotopenmarkiertem D_3 - ^{13}C -dT entwickelt. Die Markierung gelangt über die Fütterung von D_3 - ^{13}C -Methionin nach dessen Metabolisierung zu SAM im Rahmen der Methylierung von Cytosin zu D_3 - ^{13}C -5mdC in die DNA. Nach einer Desaminierung wäre folglich D_3 - ^{13}C -dT in der DNA zu finden. Mit dieser Methode könnte durch eine Bestimmung der relativen Mengen an D_3 - ^{13}C -5mdC ein deutlicher Hinweis auf Replikations- sowie auch Tet und damit

oxidationsunabhängige Demethylierung gewonnen werden. (Abb. 3A) Zudem konnte durch eine Quantifizierung von D_3 - ^{13}C -dT in verschiedenen Dnmt *Knockouts*, Dnmt3B als mögliche Desaminase identifiziert werden. (Abb. 3B)

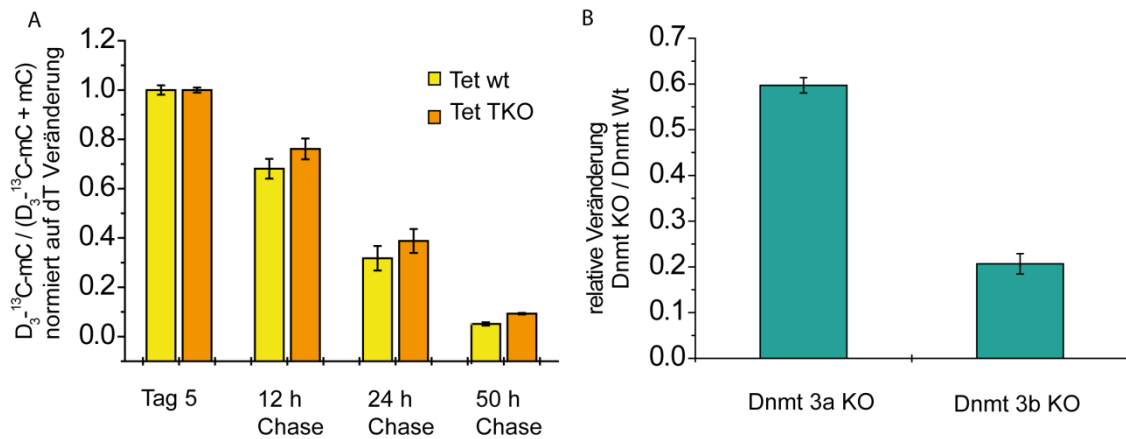


Abbildung 3: Relative Darstellung der Veränderung markierter 5mdC Spezies in An- und Abwesenheit der Tet-Enzyme (A) und Einfluss der DNA-Methyltransferasen Dnmt 3B auf den Umsatz von 5mdC zu dT (B).

Somit könnte die Desaminierung als Backup- Weg fungieren, wenn die Oxidativen Wege durch ein Fehlen der Tet Enzyme nicht mehr möglich sind.

Im weiteren Verlauf der Doktorarbeit wurde zudem eine sensitive Quantifizierungsmethode zu Bestimmung geringer Mengen von Ethenoaddukten in DNA entwickelt. In Zusammenarbeit mit der Gruppe von Prof. Dr. med. Jürgen Schlegel der Neuropathologie der TU München konnte ein Zusammenhang zwischen der Gabe eines Chemotherapeutikums und der Entstehung von Ethenoaddukten hergestellt, jedoch letztlich nicht eindeutig bewiesen werden.

Im letzten Teil der Doktorarbeit wurde die biologische Funktion der der RNA Modifikation ms^2i^6A außerhalb der tRNA näher beleuchtet. In einem SILAC-Pulldown konnten zahlreiche spezifische Binderproteine von ms^2i^6A identifiziert werden. Eine Analyse der biologischen Funktion der Binderproteine ergab einen ersten Hinweis auf einen Zusammenhang der Modifikation mit dem Vorgang des Spleißens. Durch Proteom- und Transkriptomanalysen einer Mutante ohne ms^2i^6A konnten weitere Hinweise auf eine Funktion der Modifikation im Vorgang des Spleißens (Abb.4A) sowie einige alternativ gespleißte Proteine selbst identifiziert werden. Die Anwesenheit von ms^2i^6A scheint sich somit indirekt auf die Synthese einiger Proteine, wie den mitochondrialen Transkriptionsfaktor TFAM, auswirken. Eine Auswirkung auf mitochondrielle Proteine

konnte auch in der Proteomanalyse bestätigt werden. Hier wurden mitochondriale Proteine in den *Knockout* Zellen ohne ms^2i^6A ab gereichert. (Abb.4B) Durch die Modifikation könnten somit spezielle Formen des Spleißens getriggert werden.

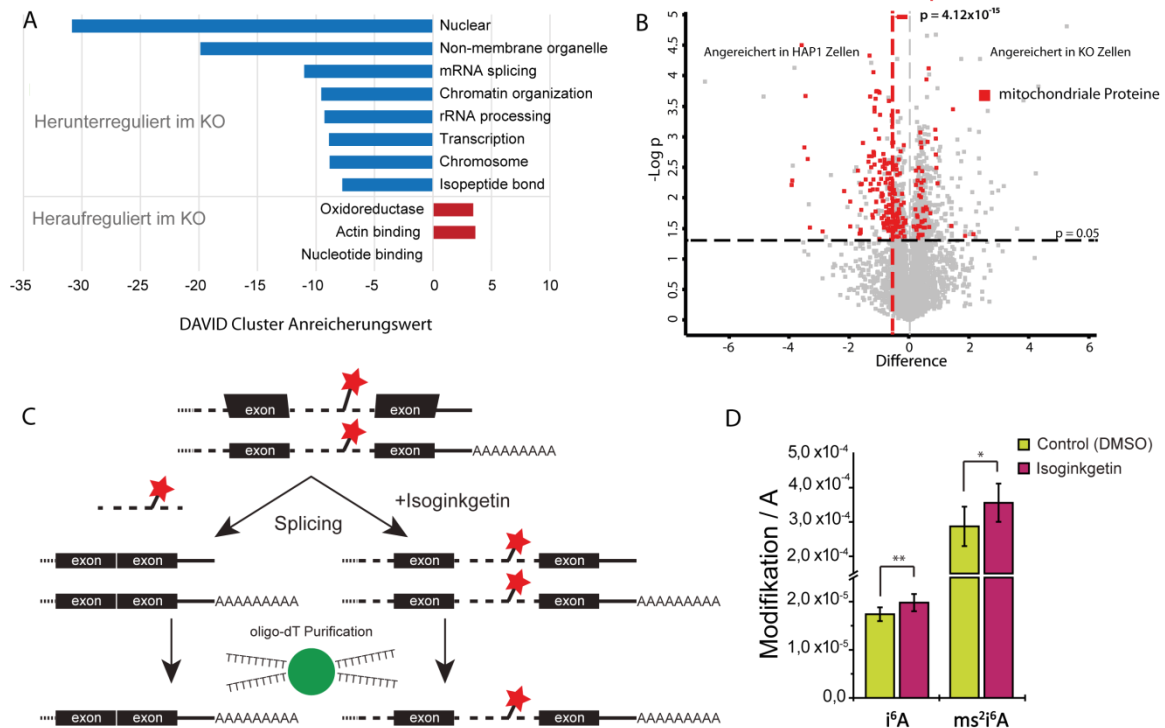


Abbildung 4: GO-Analyse der RNA-Sequenzierungs Daten von Hap Wildtyp und CDK5RAP1 KO Zellen (A). Anreicherung mitochondrialer Proteine in der Proteomanalyse der Wildtyp Zelllinie im Vergleich zum Knockout (B). Schematische Darstellung der Wirkungsweise des Spleiß-Inhibitor Isoginkgetin (C). Auswirkungen des Inhibitors auf die Mengen an i^6A und ms^2i^6A (D).

Mit Hilfe eines des Spleiß-Inhibitors Isoginkgetin konnten zudem Hinweise auf eine Lokalisierung von ms^2i^6A in den Introns oder *skipped* Exons der pre-mRNA gewonnen werden. (Abb.4C+D) Somit scheint sich ms^2i^6A in den Introns der pre-mRNA zu befinden und hier das alternative Spleißen spezieller, beispielsweise mitochondrialer, Proteine zu beeinflussen.

1 Einleitung

1.1 Modifikationen von DNA

Unser Grundverständnis über den Bauplan jedes Lebewesens, das in der Abfolge der vier DNA-Basen 2'-Desoxyadenosin (dA), 2'-Desoxyguanosin (dG), 2'-Desoxythymidin (dT) und 2'-Desoxycytidin (dC) in der Doppelhelix lange Zeit festgeschrieben zu sein schien,^[1, 2] wurde durch die Entdeckung von spezifischen Modifikationen der Bausteine stark verändert.^[3-6] Der Informationsgehalt der DNA Bausteine, festgelegt in der sogenannten DNA-Sequenz, wird durch eine gezielte Modifikation nicht verändert. Allerdings kann die Art, in der die genetische Information der DNA genutzt wird, durch DNA-Modifikationen gezielt variiert werden. Bei Änderungen, mit dem Zweck der Regulation der Transkription, spricht man von epigenetischem Code. Eine wichtige DNA-Modifikation stellt die DNA-Methylierung dar, welche Einfluss auf vielfältige biologische Prozesse hat.^[7-10]

Weitere DNA-Modifikationen sind beispielsweise Strukturänderungen der DNA (Z-DNA oder *supercoiled* DNA) und Phosphorylierungen an den Enden der DNA.

Auch bei DNA Schäden spricht man von einer Modifikation der DNA. In der Zelle ist die DNA vielen exogenen und endogenen Angriffen ausgesetzt. Das reaktive Milieu in der Zelle selbst, die Gegenwart vieler toxischer Substanzen aus der Umwelt sowie ultraviolette und ionisierende Strahlen verändern die DNA ständig. Dadurch kann es zur Modifikation oder dem Verlust von Basen kommen. Jeder DNA-Schaden muss repariert werden, wenn die genetische Information erhalten bleiben soll.

1.1.1 Epigenetik

Alle Zellen eines Organismus besitzen dieselbe genetische Information, erfüllen jedoch ganz unterschiedliche Funktionen. Dies ist nur möglich wenn spezifische Gene in den jeweiligen Zellen zu bestimmten Zeitpunkten an- oder ausgeschaltet werden. Es bedarf also eines Mechanismus der Kontrolle der Genexpression oder –supression und einer Vererbung dieser Genaktivitätsmuster an die Tochterzellen.

Eine Möglichkeit der gezielten epigenetischen Kontrolle ist die Modifikationen von Chromatin-Proteinen. Hierbei können die N-terminalen Enden von Histonen durch Methylierung, Acetylierung, Phosphorylierung, Ubiquitinierung, Sumoylierung oder ADP-Ribosylierung modifiziert werden (Abb. 1.1).^[11-13] Es konnte gezeigt werden, dass die Methylierung von Histonen an Lysin- oder Argininresten ein Signal für die Inaktivierung von Genen ist.^[14] Die Acetylierung definierter Lysinreste ist für Transkriptionsaktivierung, aber auch für die Packung von DNA in Chromatin von Bedeutung.^[15, 16]

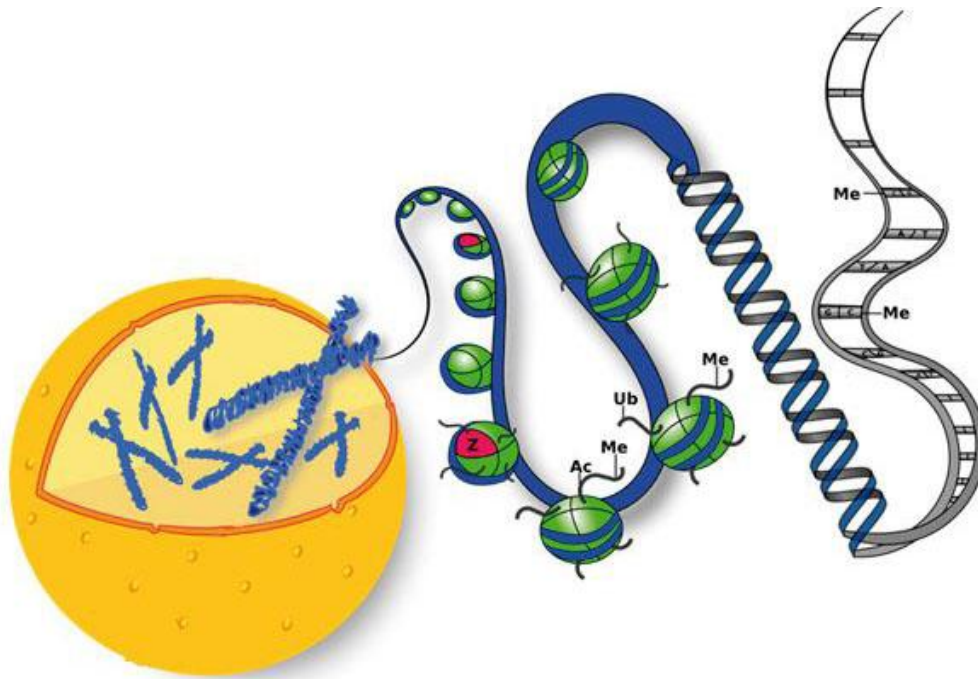


Abb. 1.1: Epigenetische Kontrolle der Zelle über verschiedene Modifikationen von Chromatin Proteinen und Methylierung der Cytosine in der DNA

Eine zweite Möglichkeit der epigenetischen Kontrolle stellt die Methylierung von Cytidin in der DNA zu 5-Methyl-2'-desoxycytidin (5mdC) dar. Diese Modifikation dient der gezielten Stilllegung von Geneabschnitten und soll im Folgenden näher beleuchtet werden.

1.1.1.1 Bildung und Häufigkeit von Methylcytosin

Die Methylierung von Cytidinen erfolgt postreplikativ durch DNA-Methyltransferasen (Dnmt). Diese erkennen bestimmte DNA-Sequenzen und übertragen eine Methylgruppe von S-Adenosyl-L-methionin (SAM) auf die C5-Position des entsprechenden Cytidins.^[17] Man unterscheidet *de novo* und *maintenance* Methyltransferasen. ^[8] Dnmt3a und Dnmt3b zählen zu den *de novo* Methyltransferasen. Sie führen Methylierungsmuster während dem Blastocystenstadium der embryonalen Entwicklung ein.^[18, 19] Diese Methylierungsmuster werden dann während der Zellteilung durch die *maintenance* Methyltransferase Dnmt1 aufrechterhalten. Dnmt1 bevorzugt hemimethylierte DNA als Substrat.^[20, 21] In somatischen Zellen erfolgt die Methylierung fast ausschließlich in CpG-Dinukleotiden. Lediglich in sehr CpG armen Genomabschnitten kommt es zur Methylierung von CpX (X= A, C oder T). Die Methylierung dieser Dinukleotide ist in Säugetieren eher selten.^[22] Ausnahmen bilden embryonale Stammzellen, Oozyten und adulte Neuronen, hier finden sich bis zu 53% dieser CpX methylierten Dinukleotide.^[23-25]

DNA Abschnitte, die sehr viele CpG-Dinukleotide aufweisen, werden CpG-Inseln genannt und finden sich hauptsächlich in Promotor-Regionen. Ist ein Promotor methyliert, wird das nachfolgende Gen in der Regel nicht abgelesen. Somit spielt der Methylierungsgrad der DNA eine wichtige Rolle in der Kontrolle der Genexpression.^[26] Bezogen auf die gesamte DNA sind in Säugetieren weniger als 4-6% aller Cytidine, jedoch 60%-80% aller CpG-Inseln methyliert.^[27, 28]

1.1.1.2 Desaminierung von Methylcytosin

Wie zuvor bereits erwähnt erfolgt die DNA Methylierung fast ausschließlich an CpG-Dinukleotiden. Da der durchschnittliche GC-Gehalt im Menschen bei etwa 40% liegt, sollte das CpG-Dinukleotid rein rechnerisch ungefähr 4% des Genomes ausmachen. Der tatsächliche Wert liegt mit 0.9% jedoch stark unter dem erwarteten Wert.^[29] Eine mögliche Erklärung für diese Diskrepanz ist eine Punktmutation von dC zu dT durch die spontane Desaminierung von 5-Methylcytidin zu Thymin.^[30, 31] Da Thymin ein natürlicher Bestandteil der DNA ist, wird diese Veränderung von den

Reparaturmechanismen zum Teil toleriert und eine dauerhafte Mutation ausgebildet. Neben der spontanen Desaminierung ist allerdings auch eine gezielte Desaminierung möglich. Dies ist bei der Demethylierung in Zygoten der Maus zu beobachten. Hier existieren zwei Phasen der DNA Demethylierung.^[32] Eine Phase findet kurz vor, die zweite Phase während der DNA Replikation statt. Zur ersten Demethylierungsphase könnte die Desaminierung von Cytosin und eine über die nukleare Uracil-DNA-Glykosylase UNG2 vermittelte *long patch* Basenexzisionsreparatur beitragen.^[32] Für die gezielte Deaminierung von 5-Methylcytidin zu Thymin sind zwei Enzyme denkbar. Zum einen könnte es sich um eine Cytosin-Desaminase zum anderen um DNA-Methyltransferasen handeln.^[33] Einige Ergebnisse deuten darauf hin, dass eine DNA Desaminierung durch die *activation-induced deaminase* (AID) in *primordial germ cells* (PGCs) und in *induced pluripotent stem cells* (iPSC) stattfindet. AID könnte auch an Kontrolle der Genexpression von Krebszellen und embryonalen Stammzellen (ES) beteiligt sein.^[34-36]

Die genauen Mechanismen dieser ersten Demethylierungsphase sind allerdings noch Gegenstand der aktuellen Forschung.^[32] Sie sollen im späteren Verlauf der Arbeit näher beleuchtet werden.

1.1.1.3 Therapeutische Relevanz von Methylcytosin

DNA Methylierung ist für eine Vielzahl von zellulären Abläufen von großer Bedeutung. Neben der Stummschaltung von Genen^[8], der Reparatur von DNA^[7] und der Chromatin Kondensation^[37], ist sie auch für das Inaktivieren von fremder DNA^[9] sowie an der Abschaltung des X-Chromosoms und von Transposomen während der Embryogenese wichtig.^[10] Somit ist klar, dass fehlerhafte DNA Methylierung zu einer Reihe von Problemen in der embryonalen Entwicklung aber auch später zu Krankheiten wie Immundefizienz, Diabetes und Krebs führen kann.^[38, 39]

Bereits 1983 konnte eine genomweite Hypomethylierung in humanen Krebszellen im Vergleich zu gesunden Zellen gezeigt werden.^[40] Da unklar war, ob dies Ursache oder Wirkung der Krebserkrankung ist, wurden in einer Studie Mäuse gentechnisch dahingehend verändert, dass sie ein schwächer ausgeprägtes Dnmt1 Allel besitzen. Dies

reduziert die Dnmt1 Expression auf etwa 10% des Wildtyps und führt zu einer genomweiten Hypomethylierung in allen Geweben. Die Mäuse sind bereits bei der Geburt sehr klein und schwach und entwickelten nach einigen Monaten aggressive T-Zell-Lymphome, welche sehr häufig eine Trisomie an Chromosom 15 aufweisen. Somit könnte die DNA Hypomethylierung durch die Destabilisierung des Chromosoms einen direkten Einfluss auf die Tumorbildung haben.^[41]

In einer anderen Studie konnte an einem Maus Modell gezeigt werden, dass die Hypomethylierung eines bestimmten Genabschnittes zum Verlust des Imprinting des Insulin Wachstumsfaktors 2 (IGF2) führt. Normalerweise ist das parentale Allel von IGF2 exprimiert, während das mütterliche Allel stumm geschaltet ist. Aufgrund seiner mitogenen und antiapoptotischen Funktionen ist IGF2 ein wichtiger autokriner Wachstumsfaktor in der Entwicklung von Tumoren.^[42, 43] Die irrtümliche Aktivierung von normalerweise stummgeschalteten mütterlichen Allelen erhöht, durch Mutationen im *Adenomatous-polyposis-coli* (APC)-Gen, das Risiko an Krebs, in diesem Fall Darmkrebs, zu erkranken.^[44]

Aber nicht nur die Hypomethylierung eines bestimmten Genabschnittes kann zu Erkrankungen führen. Auch eine Hypermethylierung, also eine zu starke Methylierung, an normalerweise unmethylierten CpG-Inseln in den Promotoren von Tumorsuppressorgenen hat die Stummschaltung des nachfolgenden Genes zur Folge. Wird ein wichtiges Gen abgeschaltet, unterbricht dies die normalen Zellfunktionen und kann ebenfalls zur Bildung von Krebs führen. In vielen Krebsarten sind die Methyltransferasen Dnmt1 und Dnmt3 sogar überexprimiert.^[45, 46]

1.1.1.4 Bildung, Häufigkeit und Verteilung von Hydroxymethylcytosin

Im Jahr 2009 konnte neben 5-Methyl-2'-desoxycytidin das oxidierte 5-Hydroxymethyl-2'-desoxycytidin (5hmdC) als weiteres Nukleosid im Genom höherer Organismen nachgewiesen werden.^[4, 47] Dabei handelt sich jedoch keineswegs um ein neuentdecktes Nukleosid. 5-Hydroxymethyl-2'-desoxycytidin wurde bereits 1952 in Bakteriophagen DNA gefunden.^[48] Mittlerweile ist 5-Hydroxymethyl-2'-desoxycytidin als „sechste Base“ des Genoms jedoch auch in höheren Lebewesen anerkannt.^[49]

Rao et al. konnten 2009 zeigen, dass hmC postreplikativ durch Oxidation von 5mdC durch Tet (ten-eleven-translocation) Enzyme gebildet wird.^[4] Bei den Tet Enzymen handelt es sich um Fe(II)- und 2-Oxoglutarat abhängige Dioxygenasen, die molekularen Sauerstoff nutzen, um Alkylgruppen über einen Radikalmechanismus zu Hydroxymethylgruppen umzuwandeln.^[50] Es konnte gezeigt werden, dass Tet-Enzyme 5mdC sowohl *in vitro* als auch *in vivo* zu hmC oxidieren können.^[51]

Bereits in den frühen 1970er Jahren wurde 5hmdC von *Penn et al.* mittels Papierchromatographie in Säugetier-DNA in unerwartet hohen Mengen gefunden. Es wurde berichtet, dass etwa 15% aller Cytosine im Mäusegehirn hydroxymethyliert sind.^[52] Diese Daten konnten jedoch nicht reproduziert werden. Mit Hilfe von radioaktiver Dünnschichtchromatographie und Massenspektrometrie (MS) konnte 5hmdC im Jahr 2009 von *Kriaucionis* und *Heintz* in Purkinje Neuronen von Säugetieren schon gefunden werden. Das Nukleosid macht in Purkinje Neuronen etwa 0.6% aller Nukleoside aus.^[47] Auch in embryonalen Stammzellen sind 0.1% aller Cytidine und damit etwa 0.03% aller Nukleoside hydroxymethyliert. Zudem konnte im Gehirn von Säugetieren eine größere Menge 5hmdC gefunden werden als in anderen Organen.

Im Jahr 2010 wurde in der Carell Gruppe eine sensitive Methode vorgestellt welche auf der Kopplung von Hochdruck-Flüssigkeitschromatographie mit Massenspektrometrie (HPLC-MS) und einer Quantifizierung durch Isotope basiert.^[53, 54] Nun konnten die Mengen an 5mdC und 5hmdC in verschiedenen Organen der Maus präzise bestimmt werden. Während bei 5mdC kaum ein Unterschied zwischen den verschiedenen Organen sichtbar ist, findet man für 5hmdC in Übereinstimmung mit *Kriaucionis* und *Heintz* eine starke Häufung von 5hmdC im Bereich des zentralen Nervensystems (ZNS). In Organen wie der Niere oder dem Herzen finden sich mittlere Werte von etwa 0.16% pro dG. Diese liegen unter den Werten im Gehirn (0.5- 0.7%), jedoch über den Werten

für die Organe Leber oder Hoden mit durchschnittlich 0.03- 0.06% pro dG. (Abb 1.2)

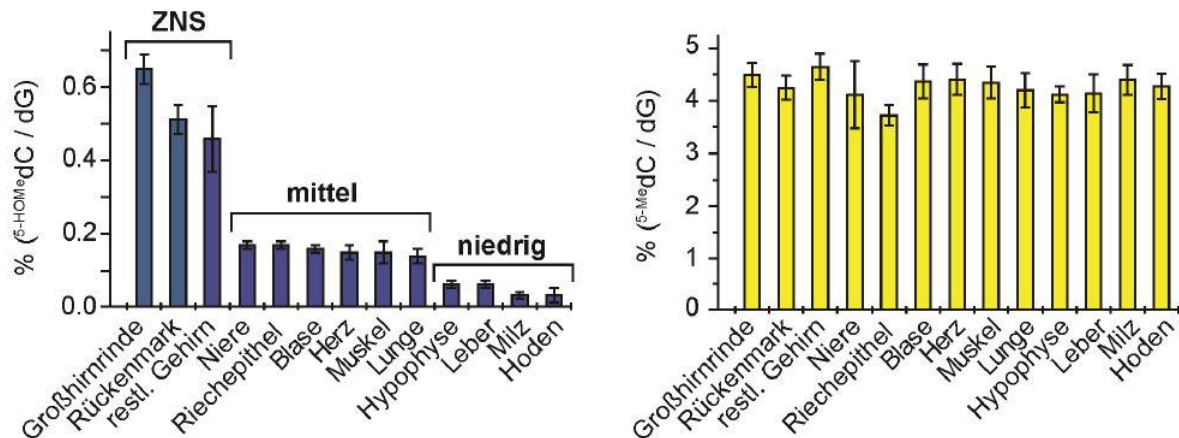


Abb. 1.2: Globale Verteilung von 5hmdC (links) und 5mdC (rechts) in verschiedenen Mäusegeweben.^[55]

Bei genauerer Betrachtung der verschiedenen Gehirnareale fällt eine hohe Variabilität auf. So finden sich in Regionen wie der Retina oder im Kleinhirn nur etwa 0.3% 5hmdC, während im Cortex oder dem Hypothalamus bis zu 0.7% pro dG gefunden werden konnten.^[25, 53, 54]

Desweiteren wurden immunohistologische Färbungen von Gewebeschnitten zur Visualisierung 5hmdC-positiver Zellen durchgeführt. Im Vergleich zu umliegenden Zellen wie Gliazellen oder neuronalen Stammzellen wiesen Neuronen die höchste Färbungsintensität auf. Daher könnte der 5hmdC-Gehalt in Neuronen noch höher sein, als es die Gewebe-Mittelwerte der LC-MS-Messungen zeigen.^[53] In einem Gewebeschnitt des humanen Cortex wurde mit 1.17% pro dG der bisher höchste genomische 5hmdC Wert gemessen. Dieser ist fast doppelt so hoch wie in der cerebralen weißen Substanz des menschlichen Gehirnes und macht etwa 20% aller DNA Modifizierungen im humanen Cortex aus.^[56] Die großen Unterschiede zwischen den Geweben und der hohe Anteil von 5hmdC in neuronalen Geweben könnten auf eine Aufgabe der Base 5hmdC in der Zelltyp-spezifischen Transkriptionsregulation bzw. auf eine neuronale Funktion hindeuten.

In einer weiteren Studie konnte zudem gezeigt werden, dass die Menge an 5hmdC während der Gehirnentwicklung beim Menschen zunimmt (Abb.1.3).

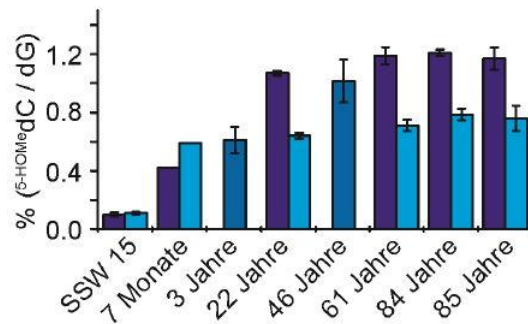


Abb. 1.3: Level an 5hmdC während der Gehirnentwicklung von Menschen. Die 5hmdC Mengen der grauen Gehirnschicht sind dunkelblau dargestellt, die der weißen in einem helleren blau.^[55]

So zeigt sich während der ersten 20 Lebensjahre ein Anstieg der Menge an 5hmdC auf das Sechsfache. Danach bleibt die Menge konstant. Dies könnte auf eine Rolle von 5hmdC während der Reifung des Gehirnes hindeuten.

1.1.1.5 Bildung und Häufigkeit der höher oxidierten Methylcytosinvarianten

Der Fund zweier weiterer oxidierter Methylcytosinvarianten im Jahr 2011 belebte die Diskussion um die Rolle von 5-Hydroxymethyl-2'-desoxycytidin neu. 5-Formyl-2'-desoxycytidin (5fdC) konnte *in vivo* von der Carell-Gruppe in embryonaler Stammzell-DNA gefunden werden.^[5] Die Modifikation wurde zu 0.02% pro dG in Stammzellen detektiert.

5-Carboxy-2'-desoxycytidin (5cadC) wurde kurze Zeit später ebenfalls in Stammzell-DNA gefunden.^[57] Beide Modifikationen werden mit Hilfe von Tet-Enzymen durch Oxidation gebildet. Obwohl Tet2 5hmdC als Substrat bevorzugt, vermag es auch 5hmdC zu 5fdC und weiter zu 5cadC zu oxidieren.^[6, 57] Anders als bei 5hmdC konnte für 5fdC und 5cadC keine Anreicherung im Gehirn gefunden werden und die Werte lagen um das 10- bis 1000-fache unter den gemessenen 5hmdC-Werten.^[58] Zudem konnte gezeigt werden, dass die TDG-Glykosylase 5fdC und 5cadC spezifisch erkennt und entfernt.^[57] Somit könnten 5fdC und 5cadC die lange gesuchten Intermediate eines aktiven Demethylierungspfadesein. Darauf soll genauer in Kapitel 1.1.1.7 eingegangen werden.

1.1.1.6 Rolle der oxidierten Modifikationen in Krankheiten

2'-Desoxy-5-hydroxymethylcytidin und die TET-Enzyme scheinen auch eine Rolle bei der Bildung von Krebs zu spielen. In vielen myeloischen malignen Erkrankungen wie der akuten myeloischen Leukämie (AML) oder dem myelodysplastischen Syndrom (MDS) wurde eine Mutation von Tet2 gefunden, welche die Fähigkeit der Zellen verringert 5hmdC zu bilden.^[59-62] So konnte in Knochenmarksproben von Patienten, welche an einer Mutation von Tet2 leiden, deutlich verringerte Level an 5hmdC detektiert werden als in gesunden Patienten.^[63] Zudem konnte gezeigt werden, dass ein erhöhter 5hmdC-Gehalt im Knochenmark keine Auswirkung auf die DNA Methylierung hatte, während ein zu niedriger 5hmdC-Gehalt sich als Hypomethylierung an der Mehrzahl der methylierten CpG Stellen zeigte. Somit scheint Tet2 für eine normale Myelopoese von großer Bedeutung zu sein und eine verminderte Entstehung von 5hmdC mit der

Myeloidtumorgenese in enger Verbindung zu stehen.^[63]

Mutationen, welche die Tet Aktivität reduzieren, können entweder direkt an den Tet Enzymen vorhanden sein, oder es sind anderen Biosyntheseenzyme defekt. Dies führt zu einer Akkumulation von Metaboliten welche die Tet Aktivität unterdrücken. Beide Arten von Mutationen finden sich in myeloischen malignen Erkrankungen, aber auch in soliden Tumoren.^[64-68] Daher gelten die Tet Enzyme als Tumor Suppressor Gene. Durch eine Reduktion ihrer Aktivität kommt es zu einer Erhöhung der 5mdC Level und dadurch zu einer Hypermethylierung anderer wichtiger Tumorsuppressor Gene.^[66, 68] Gleichzeitig scheint der Verlust von 5hmdC auch ein Merkmal vieler Krebsarten sein, wodurch sich eine mögliche Frühdiagnose für Krebs ergäbe. Der zugrunde liegende Mechanismus konnte allerdings noch nicht aufgeklärt werden.^[69]

Auch in der Carell Gruppe wurden die 5hmdC Level in Krebsgewebe untersucht. (Abb. 1.4)

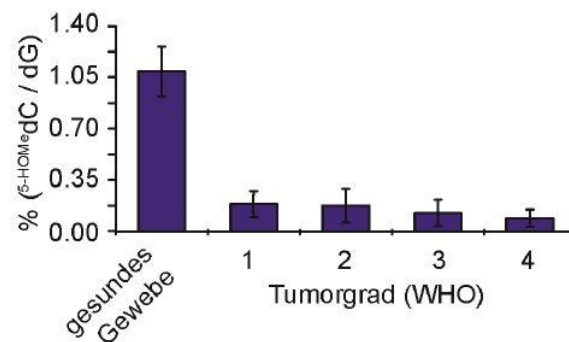


Abb. 1.4: Quantifizierung der Mengen an 5hmdC in gesunden Gewebe und Tumorgewebe. Zudem wurden unterschiedliche Tumorgrade (1-4) nach der Weltgesundheitsorganisation (WHO) untersucht.^[55]

Es konnte deutlich gezeigt werden, dass die Menge an 5hmdC in Tumorgewebe drastisch reduziert ist und mit zunehmendem Tumorgrad weiter abnimmt.

Die Frage nach dem Mechanismus soll im Folgenden näher untersucht werden. Zudem sollte beleuchtet werden inwiefern der Verlust von 5hmdC durch eine verringerte DNA Demethylierung genetisch festgelegt ist, oder ob auch die Mikroumgebung des Tumors hier eine Rolle spielt.

1.1.1.7 Embryonale Entwicklung und aktive DNA Demethylierung

In Zellen dient die Methylierung von Cytosinen in CpG Inseln dem gezielten Stilllegen von Genen. Um diese Gene wieder anzuschalten bedarf es daher eines speziellen Mechanismus diese Methylierung wieder zu entfernen. Dies kann nicht allein durch passive DNA-Demethylierung während der Replikation erfolgen. Im frühen Stadium der embryonalen Entwicklung bedarf es des gezielten An- und Ausschaltens zahlreicher Gene. Daher scheint dies der optimale Zeitpunkt zu sein, um nach aktiver DNA Demethylierung zu suchen. Nachdem eine Eizelle befruchtet wurde, spricht man von einer Zygote. Diese teilt sich zunächst durch Furchung und differenziert sich schrittweise über die 2-Zell-, 4-Zell-, 8-Zell-Stadien, zur Morula und Blastozyste weiter. (Abb 1.5)

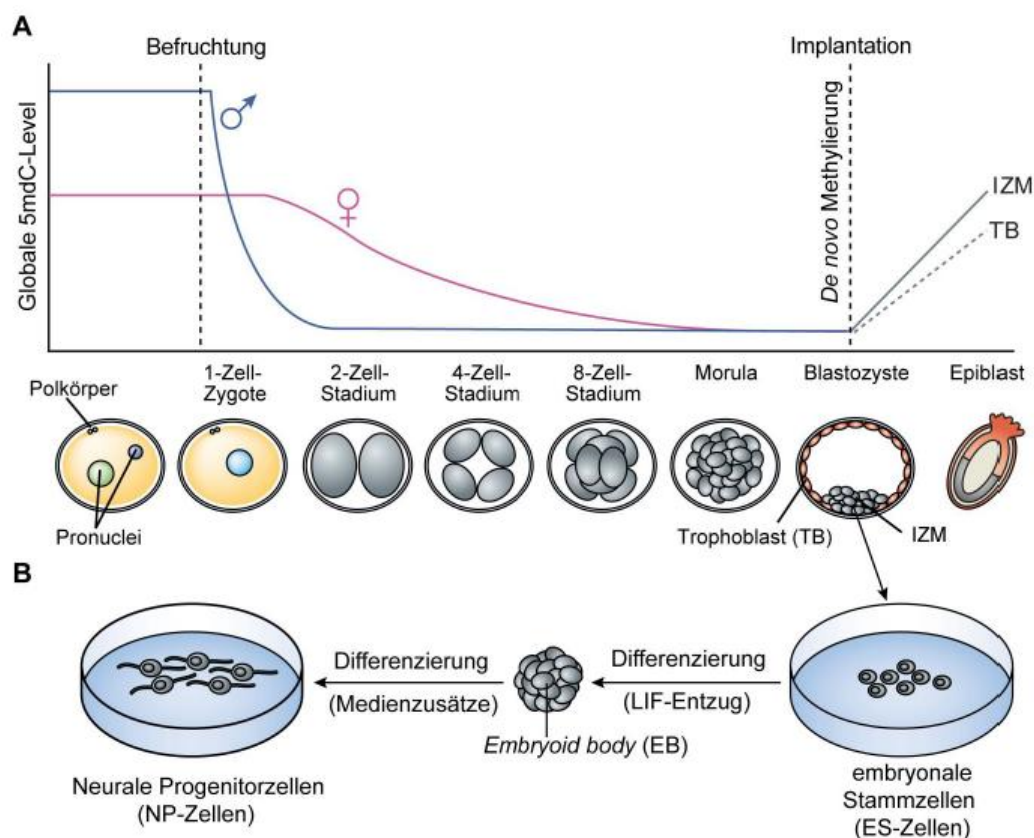


Abb.1.5: (A) Dynamik der Methylierungslevel während der frühen Embryogenese; (B) Die Gewinnung von embryonalen Stammzellen (ES-Zellen) erfolgt aus der inneren Zellmasse (IZM) im frühen Blastozysten-Stadium. Diese können unter Anwesenheit des Pluripotenz-erhaltenden Faktors LIF (*leukemia inhibitory factor*) kultiviert werden. Unter LIF-Entzug differenzieren ES-Zellen *in vitro* zu sogenannten *embryoid bodies* aus, welche in speziellen Medien in viele Zelltypen ausdifferenziert werden können.^[70-72]

Die Blastozyste besteht aus einer Hüllschicht aus Trophoblasten (TB) und der inneren Zellmasse (IZM). Die embryonalen Stammzellen (ES-Zellen) der inneren Zellmasse sind nun nicht mehr totipotent wie die Zygote, sondern nur noch pluripotent.^[73, 74] Dies bedeutet, dass sie sich zwar noch in alle Zellen eines Organismus zu entwickeln vermögen, aber keinen ganzen Organismus mehr erschaffen können. Entnommene und kultivierte embryonale Stammzellen können dank spezieller Medien im pluripotenten Status gehalten oder in unterschiedliche Entwicklungsstadien versetzt und in viele Zelltypen ausdifferenziert werden.

Betrachtet man die Methylierungslevel während der frühen Entwicklungsphase so zeigt sich kurz nach der Befruchtung bei der 1-Zell-Zygote ein massiver Verlust an 5mdC im väterlichen Genom.^[75-78] Da zu diesem Zeitpunkt noch keine Zellteilung stattgefunden hat, kann diese Abnahme nicht durch Verdünnung erklärt werden. Im mütterlichen Genom wurde lange Zeit von einer passiven, durch Replikation hervorgerufenen Demethylierung ausgegangen^[76, 77, 79], es gibt jedoch auch Hinweise auf einen aktiven Prozess.^[80]

So konnte die Gruppe von Guoliang Xu zeigen, dass nicht nur das paternale, sondern auch das maternale Genom in Zygoten eine weitreichende aktive Demethylierung vor der ersten mitotischen Teilung erfährt. Um den Anteil an aktiver Demethylierung zu bestimmen wurde die passive Demethylierung durch den Replikationsinhibitor Aphidicolin geblockt. Mit Hilfe einer Deletion von Tet3 und TDG konnte herausgefunden werden, dass die aktive Demethylierung zwar von Tet3 abhängig ist, von TDG allerdings nicht beeinflusst wird. Dies würde für einen Tet3 abhängigen, jedoch TDG unabhängigen aktiven Demethylierungsweg sprechen.^[81]

Die Gruppe von Yi Zhang konnte kurze Zeit später mit Hilfe von Tet3 reduzierten und in ihrer DNA-Replikation inhibierten Zygoten genomweite DNA Methylierungskarten vom paternalen und maternalen Genom erstellen. Sowohl im paternalen, als auch im maternalen Genom konnte gezeigt werden, dass die Inhibierung der DNA Replikation die DNA Demethylierung unabhängig von der Funktion von Tet3 unterbindet. Tet3 scheint aber die DNA Demethylierung durch Kupplung an die DNA Replikation zu erleichtern. Somit macht die replikationsabhängige Verdünnung den Hauptbestandteil der DNA Demethylierung aus. Das Enzym Tet3 scheint aber an bestimmten Orten auch eine wichtige Rolle zu spielen.^[82] Zudem konnte Yi Zhang zeigen, dass es in präimplantierten Embryos neben Tet3 abhängiger Demethylierung noch andere aktive

Demethylierungswege geben muss.^[83]

Somit ist klar, dass sowohl im paternalen als auch im maternalen Genom der Zygote eine aktive Demethylierung stattfindet. Prinzipiell sind mehrere Mechanismen für eine solche aktive Demethylierung denkbar. (Abb1.6)

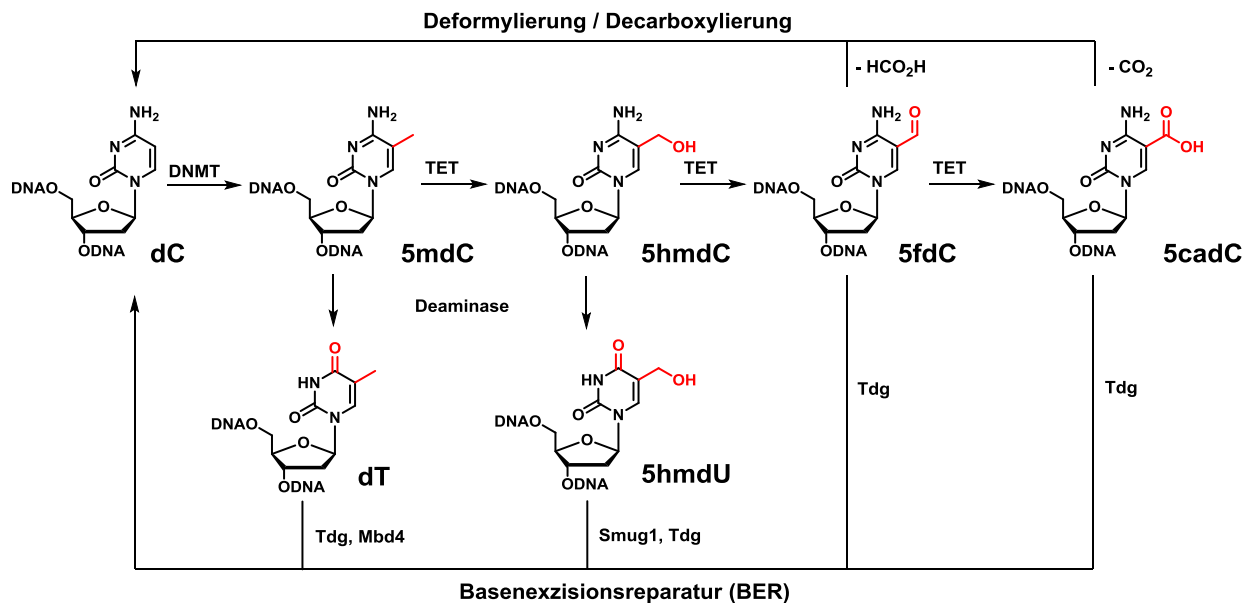


Abb. 1.6.: Denkbare Wege der aktiven Demethylierung über Oxidation und Deformylierung oder Decarboxylierung oder über Oxidation mit anschließender Basenexzisionsreparatur. Eine aktive Demethylierung von Deaminierung mit anschließender Basenexzisionsreparatur ist auch denkbar.

So sprechen einige Argumente für eine Deformylierungs- oder Decarboxylierungsreaktion der durch die Tet Enzyme gebildeten 5-Formyl- (5fdC) und 5-Carboxy-2'-desoxycytidine (5cadC).^[53, 71, 72, 84-88] Durch einen C-C-Bindungsbruch an C5 der Nukleobasen würde 2'-Desoxycytidin entstehen. Vor allem die Decarboxylierung von 5cadC ist chemisch günstig. Ähnliche Reaktionen sind von der Decarboxylierung des Orotats und Isoorotats zu Uracil bereits bekannt.^[86, 89]

Ebenso ist ein Ausschneiden der oxidierten Spezies 5fdC und 5cadC mittels TDG und anschließende Reparatur der abasischen Stellen durch die Basenexzisionsreparatur (BER) denkbar. Für viele ist dies die wahrscheinlichere Alternative, da sich hierfür zahlreiche Hinweise in der Literatur finden.^[57, 90-98] Allerdings entstehen durch die Hydrolyse der glykosidischen Bindung von 5fdC oder 5cadC abasische Stellen, welche anschließend in einem mehrstufigen BER-Mechanismus wieder repariert werden müssen. Vor kurzem konnte gezeigt werden das neben TDG auch die humanen DNA

Glykosylasen NEIL1 und NEIL2 an der Reparatur beteiligt sein könnten. So könnte TDG zunächst die abasische Stelle bilden, wonach es anschließend durch die NEILs ersetzt wird, welche dann die weitere Prozessierung des Zuckerrestes übernehmen. Dadurch könnte der Substratumsatz von TDG erhöht werden.^[99] Bei einer vermehrten aktiven Demethylierung, wie sie in der Zygote stattfindet würden nach diesem Mechanismus sehr viele abasische Stellen generiert werden. Dies könnte zu genomischer Instabilität führen, was in einer so frühen Phase der Entwicklung als kritisch angesehen wird.

Auch eine vorangehende Deaminierungsreaktion von 5mdC zu dT oder von 5hmdC zu 5hmdU mit anschließendem Ausschneiden durch eine Glycosylase und Basenexzisionsreparatur wäre prinzipiell im Rahmen der aktiven Demethylierung denkbar.^[32, 35, 36, 100-103]

Die verschiedenen Möglichkeiten einer aktiven Demethylierung sollten im Laufe der Arbeit von mehreren Seiten beleuchtet werden. Es ist denkbar, dass mehrere oder alle erwähnten aktiven Demethylierungswege in unterschiedlichen Anteilen zu einer Aufrechterhaltung der korrekten Methylierungsmuster beitragen.

1.1.2 Ethenoaddukte in DNA

1.1.2.1 Vorkommen und Bildung von Ethenoaddukten

Die Bildung von Ethenoaddukten fällt ebenfalls unter den Oberbegriff der Modifikation von DNA, allerdings handelt es sich hierbei eindeutig um eine Schädigung des Informationsträgers. Ethenoaddukte können auf zwei voneinander unabhängigen Wegen gebildet werden^[104, 105] Zum einen ist die Bildung der Addukte durch Reaktion mit Substanzen wie Vinylchlorid oder 2-Chloracetaldehyd möglich. Zum anderen können endogene Prozesse deren Bildung hervorrufen. So führt aerobe Atmung stets zur Bildung von reaktiven Sauerstoffspezies (ROS) in der Zelle. Diese begünstigen wiederum die Lipidperoxidation (LPO) und die Beschädigung mehrfach ungesättigter Fettsäuren, welche die Hauptbestandteile der Phospholipidmembranen der Zellen sind. Die durch Lipidperoxidation gebildeten Aldehyde und Hydroxylalkenale reagieren dann mit den DNA Bausteinen zu den in Abbildung 1.7 dargestellten Addukten.^[104-106]

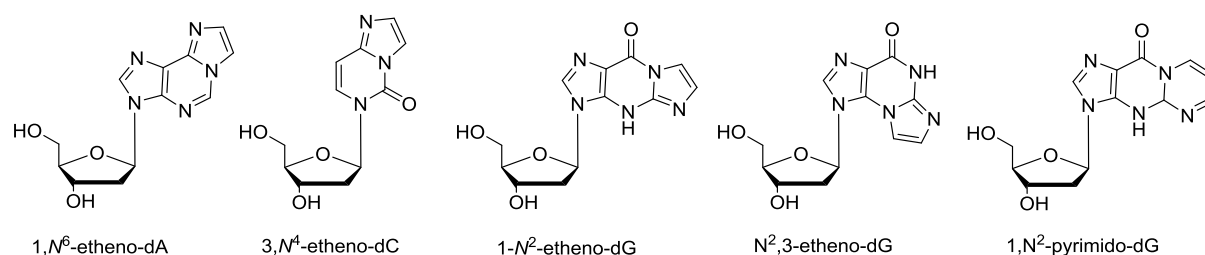


Abb 1.7.: Übersicht der möglichen Ethenoaddukte in DNA.

Gefunden wurden Ethenoaddukte in der DNA von Nagetieren und Menschen.^[107] Neben 1, N⁶-Etheno-2'-desoxyadenosin (1, N⁶-ε-dA), 3, N⁴-Etheno-2'-desoxycytidin (3, N⁴-ε-dC) und 1- N²-Etheno-2'-desoxyguanosin (1- N²-ε-dG) wurde noch zwei weitere Addukte des Guanosins gefunden: N², 3-Etheno-2'-desoxyguanosin (N², 3-ε-dG) und 1, N²-Pyrimido-2'-desoxyguanosin (1, N²-py-dG).

Mit Hilfe von DNA Glykosylasen können die Addukte innerhalb einer Basenexzisionsreparatur (BER) wieder entfernt werden. Es konnten spezielle Ethenobasen DNA-Glykosylasen gefunden werden, welche spezifisch ein bestimmtes Addukt ausschneiden können.^[106]

Es konnte zudem gezeigt werden, dass AlkB Proteine, eine Familie von Fe(II) und

α -Ketoglutarat abhängigen Dioxygenasen, in der Lage sind die hochmutagenen Ethenoaddukte aus der DNA zu entfernen.^[108, 109] Die AlkB Proteine nutzen, ähnlich den Tet Enzymen, molekularen Sauerstoff und α -Ketoglutarat um die Alkylgruppen der Ethenoaddukte zu einem Epoxid zu oxidieren. Nach darauffolgender Hydrolyse wird dieses unmittelbar als Dialdehyd abgespalten wodurch das unmodifizierte Nukleosid zurückerhalten wird.^[110, 111] (Abb.1.8)

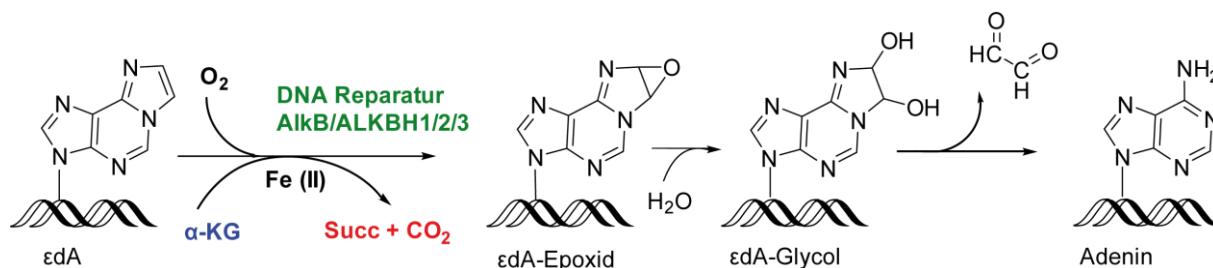


Abb.1.8: Reparatur von 1,N⁶-Etheno-2'-desoxyadenosin (1,N⁶-ε-dA) durch die Fe(II) und α -Ketoglutarat abhängigen Dioxygenasen der AlkB Familie. α -KG= α -Ketoglutarat, Succ = Succinat.

In einer Studie wurde die Reparatur der drei Ethenoaddukte 1,N⁶-Etheno-2'-desoxyadenosin (1,N⁶-ε-dA), 3,N⁴-Etheno-2'-desoxycytidin (3,N⁴-ε-dC) und 1-N²-Etheno-2'-desoxyguanosin (1-N²-ε-dG) mit 9 verschiedenen bakteriellen und 2 humanen AlkB Proteinen untersucht. Von den bakteriellen AlkB Proteinen waren die meisten in der Lage die Ethenoaddukte zu reparieren. Das humane ALKBH2 konnte alle drei Addukte reparieren, das humane ALKBH3 dagegen zeigte eine Spezifität für 3,N⁴-ε-dC in einzelsträngiger DNA.^[108]

Die Tatsache, dass Ethenoaddukte in so effizienter Art und Weise erkannt und ausgeschnitten werden legt nahe, dass das mutagene Potential der Modifikationen sehr hoch ist und die Reparatur daher von großer Bedeutung ist, um die Genstabilität aufrecht zu erhalten.

1.1.2.2 Ethenoaddukte und Krankheiten

Exozyklische Ethenoaddukte gelten tatsächlich als sehr mutagen und kanzerogen. Gebildet werden diese endogen durch die Reaktion der Purinbasen mit Lipidperoxidationsprodukten wie 4-Hydroxynonenal oder Malondialdehyd, welche durch entzündungsbedingten oxidativen Stress oder die Induktion von Cytochrom P-450 2E1 (CYP2E1) entstehen.

In der Leber wird die Bildung von CYP2E1 durch verschiedene Verbindungen wie freie Fettsäuren, Acetone und Ethanol induziert. So wurden in der Leber von Patienten, die an nicht alkoholbedingter Steatohepatitis, einer chronischen Hepatitis, aber auch in der Leber von chronischen Alkoholikern wesentlich höhere Level an CYP2E1 gefunden. Im Gegensatz zu Patienten mit chronischer Hepatitis wurden die bei Alkoholikern erhöhten CYP2E1 Level in der Leber, in den Schleimhäuten, der Speiseröhre und im Darm direkt mit der vermehrten Bildung von karzinogenen Ethenoaddukten korreliert.^[112] Bei an chronischer Hepatitis Erkrankten dagegen scheint die Bildung der Ethenoaddukte nicht hauptsächlich über CYP2E1 sondern über Entzündungsreaktionen zu verlaufen. Aufgrund der hohen Mutagenität von Ethenoaddukten wird vermutet, dass sie die alkoholvermittelte Karzinogenese durch die Induktion von CYP2E1 verursachen.^[112]

Mittlerweile kann man die Beziehung zwischen chronischen Entzündungen und Krankheiten wie Krebs nicht nur in einen losen Zusammenhang zueinander stellen, sondern tatsächlich von Ursache und Wirkung sprechen.^[113, 114]

Dennoch ist es schwierig eine Krankheit wie Krebs ganz auf Entzündungsreaktionen zurückzuführen. Es ist sehr kompliziert die Gesamtheit der chemischen und molekularen Ereignisse, die von der Entzündung zum Krebs führen, zu erfassen und zu quantifizieren.

Eine mögliche Kaskade, die von einer Entzündung zu einer Krankheit wie Krebs führt, könnte über eine lokale Infektion, Schädigung eines Gewebes oder eine Autoimmunfehlfunktion zu einer Aktivierung von Makrophagen oder neutrophiler Granulozyten/Neutrophilen verlaufen.^[115] Diese sondern dann biologisch aktive Moleküle wie Stickoxide und Superoxide im Fall von Makrophagen und hypochlorige Säure im Fall von Neutrophilen ab (Abb. 1.9).

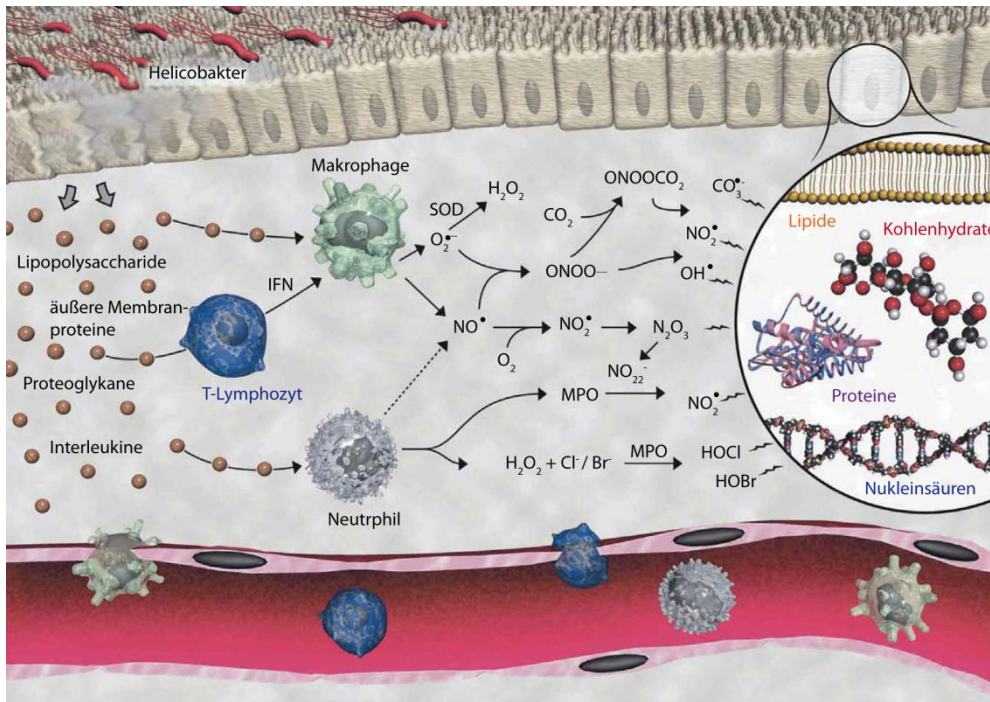


Abb. 1.9: Chemische Kaskade, die von einer Entzündungsreaktion zu Mutationen und schließlich zu Krebs führen kann.^[116]

Weitere Reaktionen mit Stickoxiden und Sauerstoff führen dann zur Bildung von N_2O_3 , NO_2 -Radikalen und Peroxynitriten. Diese reaktiven Sauerstoffspezies greifen schließlich alle möglichen Moleküle in einer Zelle an, wodurch unkontrollierte Mutationen, der Zelltod oder auch die Expression von bestimmten Onkogenen hervorgerufen werden können.^[117]

Ethenoaddukte entstehen auch auf Grund von Umwelteinflüssen. So enthalten Autoabgase, Zigarettenrauch aber auch einige Lebensmittel Acetaldehyd, welches mit den 2'-Desoxyguanosinen in der DNA zu 1- N^2 -Etheno-2'-desoxyguanosin (1- N^2 - ϵ -dG), N^2 -Ethyliden-2'-deoxyguanosin und weiter zu 1, N^2 -Propano-2'-deoxyguanosin reagiert.^[118, 119] In einer Studie wurden Ratten verschiedenen Mengen an Acetaldehyd in der Atmosphäre ausgesetzt und die Level an 1- N^2 - ϵ -dG und der beiden anderen Guanosinmodifikationen mit Hilfe von HPLC/MS bestimmt. In den Lungen der Ratten die höheren Dosen an Acetaldehyd ausgesetzt waren konnte auch höhere Level an 1, N^2 -Propano-2'-deoxyguanosin detektiert werden. Auch für das Guanosin Ethenoaddukt 1- N^2 - ϵ -dG konnte dies beobachtet werden. Während neben der Lunge auch in der Leber mehr Addukte mit steigendem Acetaldehydgehalt festgestellt werden konnten, konnte dies im Gehirn nicht beobachtet werden. Hier scheinen die Addukte nicht durch

Acetaldehyd gebildet zu werden.^[119] Dieses Wissen könnte genutzt werden, um einen Zusammenhang zwischen dem Kontakt mit Acetaldehyd und dem Auftreten von Krebs herzustellen.

1.2 RNA Modifikationen

Im Gegensatz zur DNA ist die Ribonukleinsäure (RNA) sehr viel stärker modifiziert. Als erste Modifikation wurde Pseudouridin (Ψ) bereits in den frühen 1950er Jahren entdeckt.^[120, 121] Pseudouridin ist die häufigste Modifikation in allen RNA Spezies und wird posttranskriptional durch Isomerisierung aus Uridin gebildet.^[122] Bis heute konnten über 100 Modifikationen in RNA entdeckt werden.^[123] Darunter finden sich einfache Modifikationen wie die Methylierung des Adenosins zu N^6 -Methyladenosin (m^6A) oder des Guanosins zu N^7 -Methylguanosin (m^7G). Es finden sich aber auch komplexere Modifikationen des Adenosins wie N^6 -Threonylcarbamoyladenosin (t^6A), N^6 -Isopentenyladenosin (i^6A) oder 2-Methylthio- N^6 -isopentenyladenosin (ms^2i^6A).^[123] Verbindungen wie ms^2i^6A , deren Base sehr stark modifiziert, ist bezeichnet man als hypermodifizierte Basen. Weitere Beispiele hierfür sind das Queuosin (Q) oder das Wybutosin (yW). Mit zunehmender Sensitivität der Analysemethoden konnten immer mehr und auch sehr wenig abundante Modifikationen der RNA identifiziert werden. Im Vergleich zu anderen RNAs weist die Transfer-RNA (tRNA) die höchste Modifikationsdichte wie auch am meisten verschiedene Modifikationen auf. Heute sind etwa 90 Modifikationen der tRNA bekannt.^[124, 125] In der Antikodonschleife der tRNA sind vor allem die in Abb. 1.10 markierte Position 34, die sogenannte Wobbleposition und die Position 37 (*extended anticodon*) stark modifiziert.^[126]

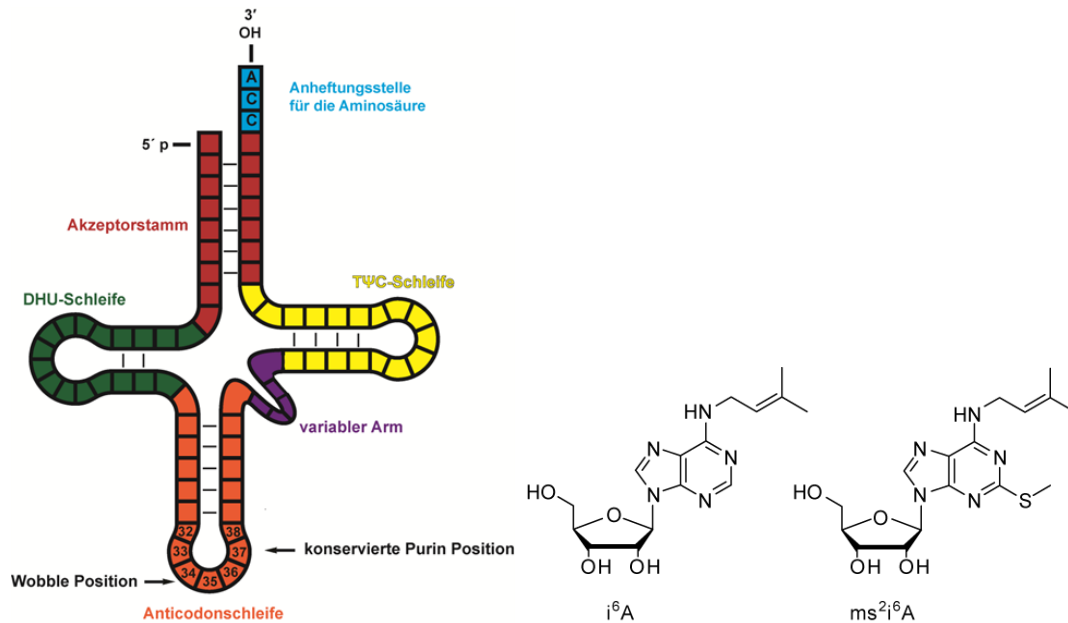


Abb. 1.10: Darstellung der zweidimensionalen Kleeblattstruktur einer tRNA und der Isopentenylmodifikationen i^6A und ms^2i^6A an Position 37.^[127]

Die Modifizierung dieser Positionen verändert die Struktur der Anticodonschleife und beeinflusst so die Erkennung der richtigen Kodon-Antikodon Bindung.^[128] An der Position 37 finden sich viele unterschiedliche Modifikationen, darunter auch die Isopentenylmodifikationen i^6A und ms^2i^6A (Abb. 1.9). Diese kommen in tRNA ausschließlich an dieser Position vor.^[129]

Im Gegensatz zu tRNA ist *messenger* RNA (mRNA) wesentlich weniger stark modifiziert. Erst kürzlich wurden neben des bekannten N^7 -Methylguanosins (m^7G) aus der *Cap*-Struktur von mRNA auch die Modifikationen Pseudouridin (Ψ), Inosin (I), 5-Methylcytosin (m^5C), N^6 -Methyladenosin (m^6A) und N^1 -Methyladenosin (m^1A) entdeckt.^[130-133] Das Vorhandensein der reversiblen Modifikation m^6A welche durch Methyltransferasen eingeführt und durch Demethylasen wieder entfernt wird hat den Begriff der RNA-Epigenetik geprägt.^[133-135] N^6 -Methyladenosin scheint an vielen Prozessen wie der mRNA Lokalisierung, dem mRNA Export, der Translation und des mRNA Spleißens beteiligt zu sein.^[135]

1.2.1 Bildung und Vorkommen der Isopentenylmodifikation

ms²i⁶A

Die Biosynthese des 2-Methylthio-*N*⁶-isopentenyladenosin (ms²i⁶A) konnte zuerst in Bakterien aufgeklärt werden. In einem ersten Schritt wird Adenosin mit Hilfe des Enzymes MiaA (Methylthio-isopentenyl-adenosine synthesis enzyme A), einer tRNA-Isopentenylpyrophosphat Transferase und Dimethylallylpyrophosphat (DMAPP) in *N*⁶-Isopentenyladenosin (i⁶A) umgewandelt. Erst im zweiten Schritt kommt es zur Thiomethylierung von i⁶A durch das Enzym MiaB (Methylthioisopentenyl-adenosine synthesis enzyme B).^[136] (Abb. 1.11)

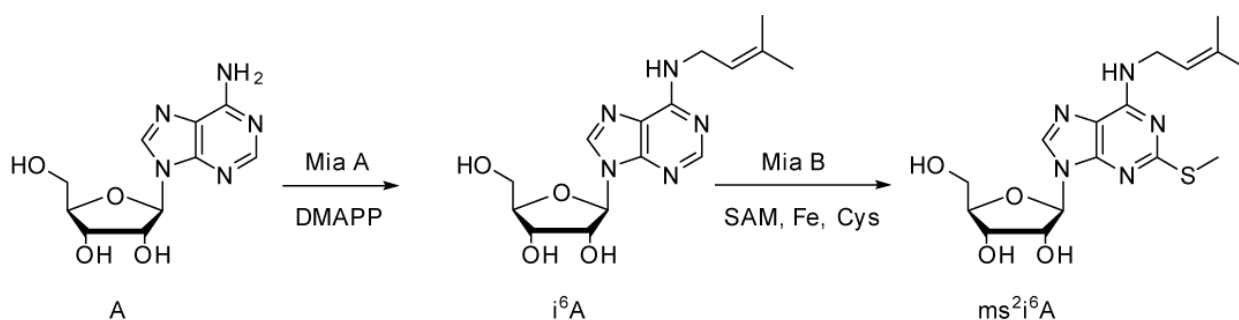


Abb. 1.11: Bakterieller Biosyntheseweg von i⁶A durch Mia A und DMAPP und ms²i⁶A durch das Enzym Mia B und SAM, Eisen und Cystein.

Das Enzym MiaA zählt zur Klasse der Isopentenyltransferasen und verwendet Dimethylallyldiphosphat (DMAPP) als Substrat für die Isopentenylgruppe. Unter Abspaltung des Diphosphates wird DMAPP mit dem Stickstoff der freien Aminogruppe an der Position 6 verbunden und damit i⁶A gebildet.^[137]

Zur Auswahl und Bindung der korrekten tRNA besitzt Mia-A eine spezielle „swinging domain“. Das Enzym erkennt die konservierte Sequenz A₃₆A₃₇A₃₈ in der Antikodonschleife^[138, 139] und dreht anschließend die zu modifizierende Base an Position 37 aus dem Antikodon in die aktive Tasche des Enzyms.^[140] Dort findet dann die Isopentenylierung statt. Mittlerweile ist auch das humane Homolog, die tRNA-isopentenyltransferase TRIT1 bekannt.^[141]

Der nächste Schritt in der bakteriellen Biosynthese von ms²i⁶A besteht in der Thiomethylierung des i⁶A durch das Enzym Mia B. Aufgrund seines CysXXXCysXXCys-Motivs in der Aminosäuresequenz wurde MiaB als Mitglied der Gruppe der Radikal-SAM Enzyme identifiziert. Es konnte gezeigt werden, dass für die Thiomethylierung Eisen III, Cystein und S-Adenosylmethionin (SAM) zwingend benötigt werden.^[142-145] Die Synthese der Thiomethylgruppe erfolgt vermutlich in zwei Schritten. Zuerst findet eine radikalische Insertion des Schwefels in die CH-Bindung statt, anschließend die Methylierung.^[146] Der genaue Mechanismus für diese Radikalreaktion ist immer noch Gegenstand der aktuellen Forschung.^[142, 146] Als humanes Homolog konnte CDK5RAP1 identifiziert werden. Nach neueren Erkenntnissen ist das humane CDK5RAP1 nicht nur ein Inhibitor von CDK5 (cyclin-dependent protein kinase 5) sondern ebenfalls in Lage i⁶A in ms²i⁶A umzuwandeln.^[147, 148] CDK5RAP1 ist ein Radikal-SAM Enzym und weist eine starke Homologie zum bakteriellen Mia B auf.^[136] In Mäusen existieren zwei Spleißvarianten des CDK5RAP1. Es konnte gezeigt werden, dass eine Spleißvariante im Cytosol, die andere im Mitochondrium lokalisiert ist.^[147] Es konnte zudem gezeigt werden, dass die cytosolische Variante, welcher am N-Terminus das mitochondrielle Importsignal fehlt, zwar dazu in der Lage war ms²i⁶A zu produzieren, also prinzipiell katalytisch aktiv ist, aber das Substrat i⁶A im Cytosol aber fehlt.^[149] Ein *Knockout* der cytosolischen Variante würde folglich nicht zu reduzierten ms²i⁶A-Level führen. Demnach scheint in Mäusen ms²i⁶A ausschließlich in Mitochondrien eine Rolle zu spielen.

Im Menschen existieren 6 unterschiedliche Spleiß-Varianten des CDK5RAP1, eine davon ebenfalls mit cytosolischer Lokalisierung.^[149] Eine Funktion außerhalb der Mitochondrien im Menschen ist daher Gegenstand der aktuellen Forschung. Dafür spricht, dass die Modifikation ms²i⁶A nicht wie bisher angenommen ausschließlich an der Position 37 in der tRNA zu finden ist. Durch die Carell Gruppe konnte gezeigt werden, dass sie in der humanen Zervix Adenokarzinom Zelllinie (HeLa) in geringen

Mengen auch in Poly-A-RNA, microRNA (miRNA) zu finden ist (Abb. 1.12).

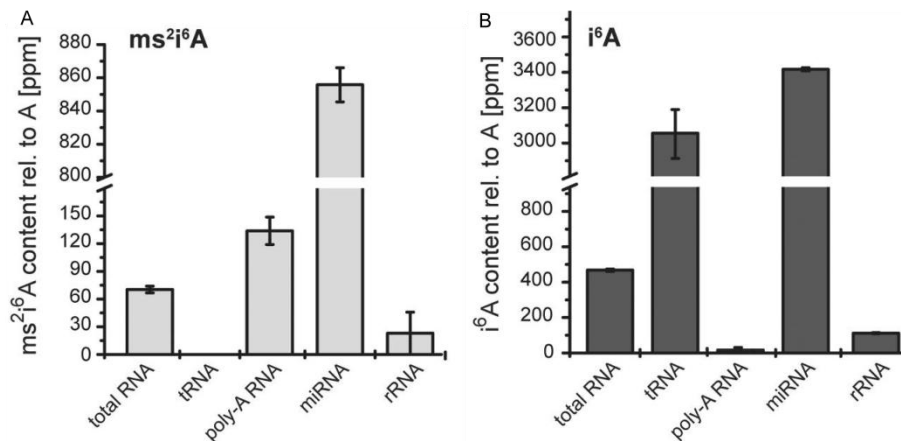


Abb. 1.12: Mengen an ms^2i^6A (A) und i^6A (B) in verschiedenen RNA Spezies aus HeLa Zellen jeweils relativ zu unmodifiziertem Adenosin dargestellt (ppm)^[147]

Auffällig ist, dass ms^2i^6A in Poly(A)-RNA detektiert werden konnte, während der Vorläufer i^6A im Vergleich zur gesamt RNA stark abgereichert ist und eher wie bisher beschrieben in tRNA zu finden ist. ms^2i^6A dagegen scheint in tRNA nicht detektierbar zu sein. In miRNA sind beide Modifikationen stark angereichert, in rRNA beide abgereichert im Vergleich zur gesamten RNA.

Die Tatsache, dass ms^2i^6A in Poly(A)-RNA angereichert ist und das Vorhandensein einer cytosolischen CDK5RAP1-Variante deuten auf eine weitere Funktion des ms^2i^6A außerhalb der tRNA und der Mitochondrien hin.

1.2.2 Funktion von ms^2i^6A

Es wurde lange Zeit davon ausgegangen, dass die hypermodifizierte Base ms^2i^6A ausschließlich an der Position 37 in tRNA zu finden ist. Hier wurde ihre Funktion im Zusammenhang mit der Translation eingehend untersucht.^[150] Es konnte bereits früh gezeigt werden, dass die Modifikation die schwache A-U-Wechselwirkung verstärkt und somit Kodon-Antikodon Wechselwirkungen unterstützt.^[151] Später zeigte sich, dass durch die Einführung der Isopentenylgruppe die Beweglichkeit der Nukleotide in der Antikodonschleife zunimmt, während sich die allgemeine Konformation der RNA nicht verändert. Man geht davon aus, dass die Destabilisierung der starren Struktur, einer

Verbesserung der Kodon-Antikodon-Wechselwirkung erzielt und alternative Strukturen besser angenommen werden können.^[152] Weitere Untersuchungen deuten darauf hin, dass ms^2i^6A in tRNA zwei konträre Aufgaben übernehmen kann. So zeigt sich eine Stabilisierung der Kodon-Antikodon-Wechselwirkung wenn die erste Basenpaarung zwischen A_{36} im Antikodon und U als erste Base im Kodon korrekt ausgebildet ist. Findet sich hier allerdings eine sogenannte „Mismatch“ Paarung, so führt die Modifikation ms^2i^6A eher zur Destabilisierung. Das Vorhandensein zweier verschiedener Kodon-Antikodon Konfigurationen unterstützt diese Theorie.^[150]

Schließlich konnte gezeigt werden, dass ein ms^2i^6A an Position 37 die Wechselwirkung von tRNA und mRNA am Ribosom an allen drei tRNA-Bindungsstellen stabilisieren und somit den Leserahmen während der Translation stabilisieren kann. Die Thiomethylgruppe unterstützt wie in Abbildung 1.13 dargestellt durch *Stacking* mit der gegenüberliegenden ersten Base des Kodons die Kodon-Antikodon Bindung während der Elongation.^[153]

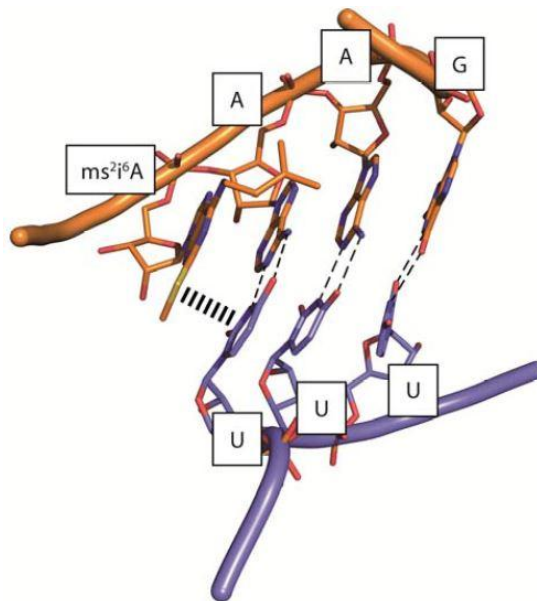


Abb. 1.13: Darstellung einer Wechselwirkung zwischen der Modifikation ms^2i^6A an Position 37 der modifizierten $tRNA^{Phe}$ (in orange dargestellt) und der mRNA (in violett dargestellt) in der ersten Phase der Elongation. Interaktionen über die Thiomethylgruppe sind durch die gestrichelte Linie dargestellt. Die Daten wurden der Kristallstruktur 3I8G entnommen.^[153]

Somit hat die hypermodifizierte Base ms^2i^6A je nach Sequenzkontext sowohl stabilisierende als auch destabilisierende Wirkungen auf die Kodon-Antikodon Wechselwirkung in der tRNA und kann durch *Stacking* den Leserahmen während der

Translation beeinflussen. Es stellt sich die Frage welche Funktionen ms²i⁶A in Poly-A-RNA ausübt, ob es neben dem Einfluss auf die Translation auch einen Einfluss auf die Proteinsynthese direkt haben könnte.

1.2.3 Rolle der Modifikation ms²i⁶A in Krankheiten

Bedingt durch die ihre Position in tRNA hat die Modifikation ms²i⁶A einen Einfluss auf die Translation. Durch stabilisierende oder destabilisierende Wechselwirkungen werden bestimmte Gene mehr oder weniger exprimiert, was auch Auswirkungen auf Krankheiten haben kann.

Bereits 1992 konnte das Fehlen der Modifikation i⁶A, des ms²i⁶A Vorläufers, in der tRNA von Bakterien mit der Stärke und Ausprägung ihrer Virulenz in Verbindung gebracht werden.^[154] Durch eine Mutation des Mia A Gens kann in *Agrobacterium tumefaciens* das für eine Infektion benötigte Virulenz Gen nicht mehr exprimiert werden.^[154] Kurze Zeit später wurde ein ähnlicher Effekt auch für die Modifikation ms²i⁶A selbst beschrieben. Im human-pathogenen Bakterium *Shigella flexneri* werden bestimmte Virulenz-Gene bei einem Fehlen der Modifikation ms²i⁶A weniger stark exprimiert und somit die Pathogenität des Bakterium deutlich abgeschwächt.^[155]

In mitochondrialen tRNAs (mt-tRNA) von Säugetieren konnten 15 unterschiedliche Modifikationen an 18 verschiedenen Positionen identifiziert werden.^[156] Viele dieser Modifikationen konnten mit der Entwicklung mitochondrialer Krankheiten wie mitochondrieller Myopathie, Enzephalopathie oder MELAS (*Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes*) und MERRF (*Myoclonic Epilepsy with Ragged Red Fibers*) in Verbindung gebracht werden.^[157] Die Modifikation ms²i⁶A konnte in Mäusen in 4 unterschiedlichen dieser mitochondrialen tRNAs (mt-tRNA^{Phe}, mt-tRNA^{Trp}, mt-tRNA^{Tyr} und mt-tRNA^{Ser(UCN)}) nachgewiesen werden.^[149, 157]

Mit Hilfe eines *Knockout* des mitochondrialen CDK5RAP1 in Mäusen konnte ein Phänotyp ermittelt werden. Der KO resultierte in einem stark reduzierten Level an ms²i⁶A in allen vier untersuchten mt-tRNAs (mt-tRNA^{Phe}, mt-tRNA^{Trp}, mt-tRNA^{Tyr} und mt-tRNA^{Ser(UCN)}) der Mäuse. Dieser niedrige Level an Modifikation führt dann zu einer Veränderung in der Translation mitochondrialer Proteine, was wiederum zu einer

Verringerung der Atmungsaktivität und einem Ansteigen der ROS (*reactive oxygen species*) in der Skelettmuskulatur und im Gewebe des Herzen führt. Dies bedingt schließlich eine verminderte oxidative Phosphorylierungsaktivität (OXPHOS) und führt somit zu einer verminderten Gewinnung von ATP. Dies kann zu Krankheiten wie Myopathie oder einer kardialen Dysfunktion führen.^[149] Dieser Zusammenhang macht deutlich, dass vor allem unter oxidativen Stress-Bedingungen eine Regulation der mitochondrialen Proteine durch die Thiomethylmodifikation ms²i⁶A besonders wichtig ist, um Krankheiten zu verhindern.

Auf Grund eines bekannten Zusammenhangs zwischen einem Mangel an CDK5RAP1 und Apoptose wurde in einer anderen Studie die Auswirkung einer Verringerung der Mengen an CDK5RAP1 auf Brustkrebs-MCF-7-Zellen untersucht. Dafür wurde CDK5RAP1 siRNA (*small interfering RNA*) in MCF-7 Zellen transfiziert. Der Mangel an CDK5RAP1 äußerte sich in einer Unterdrückung des Tumorwachstums und einem Arrest der Zellen in der G2/M Phase des Zellzyklus. Zudem zeigte sich vermehrter Zelltod und eine gesteigerte Bildung von ROS. Es konnte auch eine Hochregulation der Proteine p-JNK (*phospho-c-Jun N-terminal kinase*), p53, Caspase 9 und Caspase 3 beobachtet werden. Mit einem ROS Inhibitor wie N-Acetyl-cystein oder einem Inhibitor von der p-JNK konnten die Expression von p-JNK, p53, Caspase 9 und Caspase 3 reduziert und die Apoptose verhindert werden. Somit scheint ein Mangel an CDK5RAP1 über den ROS/JNK-Pathway zu Zellzyklusarrest und Apoptose in Brustkrebszellen führen.^[158] Ob es dabei eine Rolle spielt, dass bei einem Mangel an CDK5RAP1 auch ein Mangel an ms²i⁶A vorliegt, oder ob der Effekt auf die Brustkrebszellen unabhängig von den Mengen an ms²i⁶A in den Zellen ist, konnte in dieser Studie nicht geklärt werden. Eine Studie der Carell Gruppe konnte allerdings zeigen dass MCF-7 Zellen etwas höhere Level an ms²i⁶A aufweisen als beispielsweise HeLa Zellen.^[127] Verglichen mit der humanen mikrovaskulären Endothelzelllinie (HMEC), waren die Level an ms²i⁶A in MCF-7 Zellen sogar deutlich höher. Daher ist anzunehmen, dass nicht nur CDK5RAP1 sondern auch die Modifikation ms²i⁶A selbst eine Rolle beim Tumorwachstum von Brustkrebszellen spielt. Mit einem Mangel an CDK5RAP1 geht zudem nicht nur ein Mangel an ms²i⁶A sonder auch eine gesteigerte Menge an i⁶A einher. So führt ein Anstieg von der i⁶A Level zu Zellzyklusarrest in Brustkrebszellen, dies konnte durch die Fütterung von i⁶A belegt werden.^[158]

Somit können der komplexen Thiomethylmodifikation $\text{ms}^2\text{i}^6\text{A}$ in tRNA wichtige Funktionen sowie auch ein Zusammenhang mit einigen Krankheiten zugeschrieben werden. Daher wäre es interessant ihre Funktion in mRNA ebenfalls aufzuklären. Möglicherweise ergeben sich auch hier Zusammenhänge zu Fehlfunktionen oder Krankheiten die durch die An- oder Abwesenheit der Modifikation beeinflusst werden.

2 Aufgabenstellung

Bereits im Jahr 2009 wurde die epigenetische Modifikation 5-Hydroxymethyl-2'-desoxycytidin im Genom von Säugetieren in Stammzellen und im Gehirn entdeckt. Zwei Jahre später konnten die höher oxidierten Spezies 5-Formyl-2'-desoxycytidin und 5-Carboxy-2'-desoxycytidin entdeckt werden.

Das erste Ziel dieser Doktorarbeit bestand darin mit einer von Toni Pfaffeneder entwickelten UHPLC-MS/MS Methode zur Quantifizierung der epigenetischen Modifikation verschiedene Organe der Maus hinsichtlich ihres Modifikationsgehaltes an 5fdC und 5cadC zu untersuchen, sowie die Modifikationen während des Alterungsprozesses zu beobachten. Für eine Reihe verschiedener Kooperationsprojekte wurde die Methode so verändert, dass die Level der epigenetischen Modifikationen in verschiedenen Tumorgeweben, Leukämie Zelllinien und Retinapräparationen bestimmt werden konnten.

Die Anwesenheit der höher oxidierten Spezies 5fdC und 5cadC wirft die Frage nach deren Funktion und die Möglichkeit der aktiven Demethylierung über diese Spezies auf. So sollte die Möglichkeit einer Decarboxylierung oder Deformylierung mit Hilfe eines Isotopenverfolgungsexperimentes mit einem doppelt deuterierten und doppelt ^{15}N -markierten Cytosines ($\text{D}_2\text{-}^{15}\text{N}_2\text{-dC}$) erforscht werden. Auch hier sollte eine Methode zur Quantifizierung der verschiedenen isotoopenmarkierten Nukleoside entwickelt werden. Eine weitere Möglichkeit zur aktiven Demethylierung ergibt sich aus der Deaminierung von 5mdC zu dT. Auch hierfür sollte eine sensitive Quantifizierungsmethode entwickelt werden.

Im weiteren Verlauf der Doktorarbeit sollte auch eine Methode zur sensitiven Quantifizierung von Ethenoaddukten in DNA entwickelt werden. Diese sollte in einer Kooperation mit der Gruppe von Prof. Dr. med. Jürgen Schlegel der Neuropathologie der TU München dazu genutzt werden den Gehalt an Ethenoaddukten in bestimmten Geweben zu bestimmen.

Im letzten Teil der Doktorarbeit sollte dann die biologische Funktion der RNA Modifikation 2-Methylthio- N^6 -isopentenyladenosin ($\text{ms}^2\text{i}^6\text{A}$) untersucht werden. Hierzu wurde in einer Kooperation mit der Gruppe von Michiel Vermeulen Interaktionspartner der Modifikation $\text{ms}^2\text{i}^6\text{A}$ identifiziert werden und die biologische Funktion über diese Interaktionspartner untersucht werden. Hierzu wurden Untersuchungen des Proteoms

sowie des Transkriptoms einer Mutante ohne $ms^{2i6}A$ untersucht.

3 Veröffentlichte Arbeiten

3.1 Altersabhängige Level von 5-Methyl-, 5-Hydroxymethyl- und 5-Formylcytosin in Hirngeweben des Menschen und der Maus

Mirko Wagner,^[+] Jessica Steinbacher,^[+] Theo F. J. Kraus,^[+] Stylianos Michalakis, Benjamin Hackner, Toni Pfaffeneder, Arshan Perera, Markus Müller, Armin Giese, Hans A. Kretzschmar, Thomas Carell. (^[+] geteilte Erstautorenschaft)

Prolog

Die absoluten Level von 5-Hydroxymethyl- (5hmdC) und 5-Methylcytosin (5mdC) in humanen Hirngeweben wurden in Abhängigkeit vom Lebensalter bestimmt. Zusätzlich wurden die absoluten 5-Formylcytosin Level (5fdC) in adulten Individuen und die Level dieser modifizierten Cytosinbasen in menschlichen Neuronen quantifiziert. Wir vergleichen diese Daten mit den altersabhängigen Mengen von 5fdC, 5hmdC und 5mdC in Hirngeweben der Maus. Bei 5hmdC wird mit zunehmendem Alter zunächst ein Anstieg der Modifikationsdichte beobachtet. Im voll entwickelten Gehirn steigt der 5hmdC-Level nicht weiter an, sondern bleibt auf einem konstant hohen Niveau von ca. 1.2 % beim Menschen. Dieser Wert ist etwa doppelt so hoch wie das 5hmdC -Level im zerebralen Cortex der adulten Maus. Im Unterschied dazu sinkt der 5fdC-Level in den frühen Entwicklungsstadien mit zunehmendem Alter stark ab. Bei 5hmdC könnte es sich dementsprechend um eine stabile epigenetische Modifikation handeln, während 5fdC in den frühen Gehirnentwicklungsphasen vermutlich eher als Intermediat der aktiven DNA-Demethylierung auftritt.

Autorenbeitrag

Die Isolation der DNA, die massenspektrometrischen Quantifizierung, sowie die Auswertung und Interpretation der Daten führte ich in enger Zusammenarbeit mit *Mirko Wagner* durch. *Theo F. J. Kraus* stellte die präparierten Gewebe aus Mäusen zur Isolation von genomische DNA zur Verfügung. Für eine weitere Auflistung des Beitrags der übrigen Autoren siehe Manuskript.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags.

Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues**

Mirko Wagner, Jessica Steinbacher, Theo F. J. Kraus, Stylianos Michalakakis, Benjamin Hackner, Toni Pfaffeneder, Arshan Perera, Markus Müller, Armin Giese, Hans A. Kretzschmar†, and Thomas Carell*

Abstract: The absolute levels of 5-hydroxymethylcytosine (hmC) and 5-methylcytosine (mC) in human brain tissues at various ages were determined. Additionally, absolute levels of 5-formylcytosine (fC) in adult individuals and cytosine modification levels in sorted neurons were quantified. These data were compared with age-related fC, hmC, and mC levels in mouse brain samples. For hmC, an initial steady increase is observed, which levels off with age to a final steady-state value of 1.2 % in human brain tissue. This level is nearly twice as high as in mouse cerebral cortex. In contrast, fC declines rapidly with age during early developmental stages, thus suggesting that while hmC is a stable epigenetic mark, fC is more likely an intermediate of active DNA demethylation during early brain development. The trends in global cytosine modification dynamics during the lifespan of an organism are conserved between humans and mice and show similar patterns in different organs.

In addition to the four canonical Watson and Crick bases and 5-methylcytosine (mC), it has recently been discovered that mammalian DNA also contains 5-hydroxymethylcytosine (hmC),^[1] 5-formylcytosine (fC),^[2] and 5-carboxycytosine^[2b,3]. The biological role of these cytosine modifications is presently a focus of intense research, to which end the amount and distributions of the new bases need to be analyzed as the first step. Particularly large amounts of the base hmC were found in DNA isolated from mouse brain tissues.^[1,4] In cerebral cortex, for example, the hmC level reaches up to 0.8 % with respect to all present cytosines.^[4,5] Additionally, an increase in hmC content was observed during postnatal development in mice.^[4,6] These findings directly lead to the question of how hmC levels in humans vary over a lifetime, since knowledge of

age-dependent changes in hmC and fC levels is essential in order to gain insight into potential functions of these newly found bases.^[7] So far, however, information about cytosine modification levels in human tissues is limited. Herein we report absolute cytosine modification levels in human brain tissues at fetal and different postnatal developmental stages. We compare these data with modification levels previously reported by us for adult individuals.^[8] Quantification was performed by using our reported LC–MS-based isotope dilution method, in which synthetic isotopically labelled mC, hmC, and fC nucleosides are used as internal standards (Figure 1).^[5c,8,9]

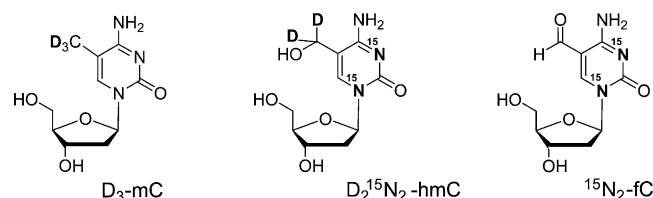


Figure 1. Depiction of the isotopically labeled mC, hmC, and fC nucleosides used as internal standards for mass spectrometry based quantification (D and ¹⁵N atoms are highlighted in bold).

For analysis, brain tissues from humans of different ages were provided by the BrainBank Munich. Tissues from a 15 week old fetus (15. WOP) and from four individuals between the ages of 0.6 and 88 years were analyzed. To this end, DNA was separately isolated from grey matter (cerebral cortex) and white matter of the cerebrum. The data obtained were compiled with previously reported data from four

[*] Dipl.-Chem. M. Wagner,^[‡] M. Sc. J. Steinbacher,^[‡] Dr. B. Hackner, Dr. T. Pfaffeneder, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstr. 5-13, 81377 Munich (Germany)
E-mail: Thomas.Carell@lmu.de

Dr. T. F. J. Kraus,^[‡] Prof. Dr. A. Giese, Prof. Dr. H. A. Kretzschmar
Center for Neuropathology and Prion Research (ZNP)
Ludwig-Maximilians-Universität München
Feodor-Lynen-Str. 28, 81377 Munich (Germany)

Priv.-Doz. Dr. S. Michalakakis, M. Sc. A. Perera
Center for Integrated Protein Science at the Department of Pharmacy - Center for Drug Research
Ludwig-Maximilians-Universität München
Butenandtstr. 5-13, 81377 Munich (Germany)

[‡] These authors contributed equally to this work.

[†] deceased

[**] We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support through the SFB program (SFB 749 and 1032). Further support was obtained from the Excellence Cluster (Center for Integrated Protein Science, CiPS^M). T.P. thanks the Fonds der Chemischen Industrie for a graduate fellowship.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201502722>.

© 2015 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

individuals aged 61 (two individuals), 84, and 85.^[8] In addition, from the 22 and 85 year old individuals, we were able to obtain DNA from cerebellum, again from both white and grey matter. Because prior studies with mouse tissues had revealed that cerebellum contains 40 % less hmC compared to cerebrum,^[4] a similar comparison in human tissues appeared highly desirable. Our quantification results regarding the absolute global levels of hmC in human brain tissues^[10] are summarized in Figure 2.

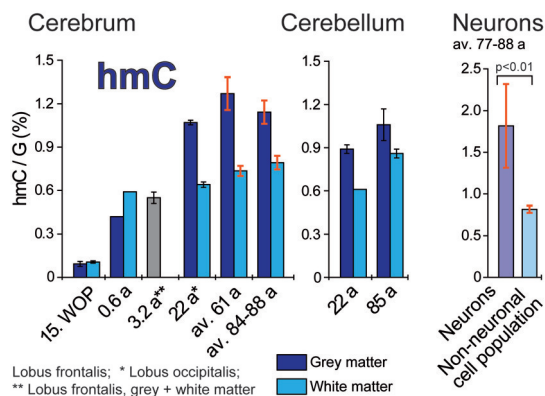


Figure 2. Age-related hmC levels in human cerebrum, cerebellum, and different brain cell populations. Values are given as modifications per 100 guanine bases. Guanine (G) was chosen as a reference because the overall G content is equal to the sum of cytosine (C) and the cytosine derivatives mC, hmC, and fC. In cerebrum and cerebellum, the white and grey matter were analyzed separately. Data for samples 61a (two individuals), 84a, and 85a (here for cerebrum only) were taken from Kraus et al.^[8] Cerebral occipital grey matter tissue of five human individuals aged 77–88 was sorted into neuronal and non-neuronal nuclei which were then analyzed separately. Black error bars represent the standard deviation of two to four technical replicates, thick red error bars represent the standard deviation of two to four biological replicates (the number of replicates was dependent on sample size). In the case of the 0.6 year old individual, owing to limited sample size, only one measurement could be performed. For statistical analysis, Student's unpaired two-tailed *t*-test was used. WOP=week of pregnancy, a=anno (year), av=average.

Our data show that in grey and white matter of the human cerebrum, the hmC level strongly increases with age. While in the cerebrum of a 15 week old fetus, only 0.1 % hmC is present in both white and grey matter, the fully developed human brain shows a constant hmC level of 1.2 % in the grey matter, which is rich in neuronal cell bodies.^[8] This is nearly twice the level detected in mouse cerebral cortex.^[4] In white matter, with myelinated axons surrounded by glial cells as the main components, hmC is less abundant. Nevertheless, the mean level of 0.75 % hmC observed in adults is higher than in most mouse brain tissues.^[4,8] Furthermore, in white matter the increase of hmC levels off already at the age of approximately one year. This is much earlier than in grey matter, where we observe a steady-state level only after the age of 22. Although white matter initially shows a higher hmC level compared to grey matter, this order is inverted during adolescence. Similar to cerebellum, the human cerebellum shows an age-dependent increase in hmC content, although at lower absolute levels. Our findings are in good agreement with mouse data available from others^[6] and our group.^[4]

We wondered whether the differences observed between the hmC levels in grey and white matter could be explained by large differences in the hmC content of neurons and glial cells. To answer this question, cerebral occipital cortex tissue samples from five individuals aged 77–88 were sorted into neuronal and non-neuronal nuclei. Cytosine modification content was then determined separately for both populations. This experiment indeed revealed a higher level of hmC in neurons compared to non-neuronal cells (1.82 % vs. 0.82 %; Figure 2).^[10d]

In order to relate our hmC data to cytosine methylation, we next quantified the levels of mC in human cerebrum and cerebellum at early developmental stages and during adolescence,^[6a, 10b–f, 11] again separately for grey and white matter. The resulting data, together with published data from four older individuals,^[8] are compiled in Figure 3.

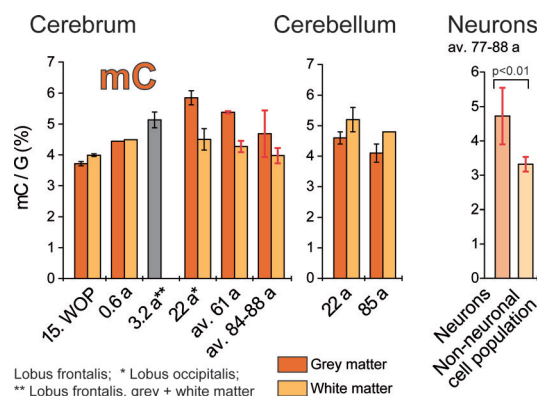


Figure 3. Age-related mC levels in human cerebrum, cerebellum, and different brain cell populations. Data for samples 61a (two individuals), 84a, and 85a (here for cerebrum only) were previously reported by us.^[8] Further details as in Figure 2.

In white matter, an age-dependent change for mC is barely detectable. Here, the mC level remains roughly constant at about 4.3 % during a human's lifetime. A moderate age effect is observed in the neuronal cell-body-rich grey matter, where mC values increase from 4.4 % (0.6a) to a maximum of roughly 6 % (22a). This finding is in full agreement with a recent study by Lister et al., who measured age-dependent *de novo* methylations in human and mouse lobus frontalis.^[6a]

In cerebellum, no age effect is observed. In the cell-sorting experiment, an elevated mC level was observed in neurons with respect to the non-neuronal cell population (Figure 3 and the Supporting Information).^[6a, 10d, 11a] In summary, our data show that global mC levels in humans change only slightly with age.

We next studied the levels of fC^[2a] in human brain tissues. This base^[2b] is proposed to be associated with active DNA demethylation.^[1b, 2, 3, 5a, 12] Because of limited human tissue availability, in combination with the fact that more tissue is needed to detect the low levels of fC,^[5c] data could only be obtained for the individuals aged 61 (sample 61*) and 85 (Figure 4).

In the cerebrum of adult human individuals, we observe fC levels of roughly $2\text{--}3 \times 10^{-4}$ % fC/G, which correspond to

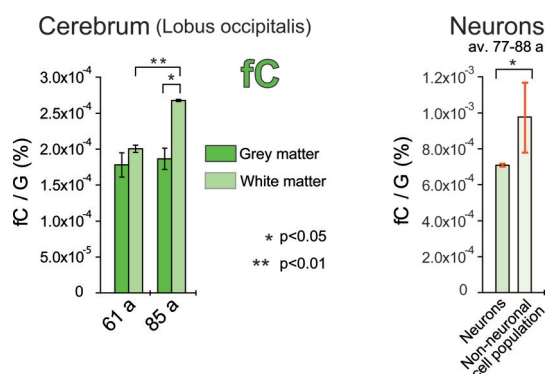


Figure 4. Global fC levels in cerebrum tissues from two human adults and in different brain cell populations. Further details as in Figure 2.

$4-8 \times 10^{-7}$ fC modifications per nucleoside (N). Similar levels of fC were found in the grey matter of the two individuals. By contrast, the fC levels differ in white matter. Here, the 85 year old individual showed an increased fC content in white compared to grey matter, while for the individual aged 61, comparable levels of fC were detected in both tissues. Consistent with this, the cell-sorting experiment revealed that with 9.8×10^{-4} % fC/G, more fC is present in the non-neuronal cell population, while in neurons, we observed a fC level of only 7.1×10^{-4} % (see Figure 4 and the Supporting Information). It should be noted that the fC levels observed by us in human cerebrum are as high as the level of fC in adult mouse cerebrum (see also Figure 5).^[5c] For a comparison of fC and hmC levels reported in this study with previously published data regarding the levels of these bases in mouse embryonic stem cells and organs at postnatal day 90, see Figures S1 and S2 in the Supporting Information.

To further study the age-dependent changes of cytosine modification levels in the mammalian brain and to overcome the limited human tissue availability, we expanded our investigations to mouse tissues. We analyzed fC, hmC,^[4,6,13] and mC^[6a,g,11c] levels in cerebrum tissue (cerebral cortex) sampled at postnatal day 1 (p1, newborn) and day 14 (p14), and from adult mice at postnatal day 90 (p90), postnatal month 12 (pm12), and postnatal month 18 (pm18; Figure 5).

Besides the expected increase in hmC content with age, we surprisingly observe a strong decrease in the fC modification level at early developmental stages.^[9b] Because fC is proposed to be generated from hmC by Tet-induced oxidation,^[2b] the decreasing fC level does not fit with the increasing amounts of hmC during brain development. One possible explanation for this counterintuitive observation is that fC is indeed an intermediate of an active DNA demethylation process^[11b,2,3,5a,12] and that active demethylation events in the brain cease with age. The higher fC levels in early lifetime suggest a faster hmC turnover via oxidation to fC,^[2b] followed by active demethylation.^[2,3,5a,12d,e] With increasing age, these as yet ill-defined processes may slow down, which would lead to the observed accumulation of hmC.

We observe similar trends in mouse kidney tissue, which was previously found to contain intermediate levels of hmC.^[5a] Here too, a low hmC level in young animals

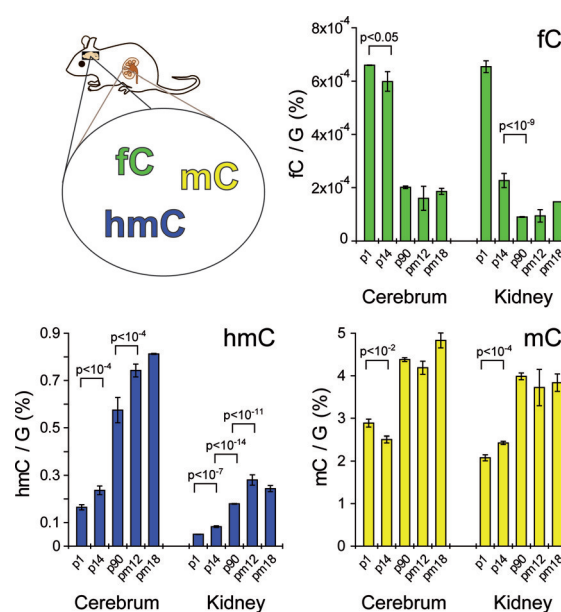


Figure 5. Global levels of fC, hmC, and mC in mouse cerebrum (cerebral cortex) and kidney at postnatal days 1 (p1) and 14 (p14) and in adult mice aged 90 days, 12 months, and 18 months. Values are given as modifications per 100 guanine bases. Guanine (G) was chosen as reference because it amounts to the sum of cytosine (C) and its derivatives mC, hmC, and fC. Samples were taken from wild-type mice (C57-BL6). Error bars represent the standard deviation of two (p1) or three (p14–pm18) biological replicates. For statistical analysis, Student's unpaired two-tailed *t*-test was used. pm = postnatal month.

corresponds to a high fC level, while higher hmC levels during later developmental stages go hand in hand with low fC values. This observation is in good agreement with the idea that reduced active demethylation is a reason for hmC accumulation in tissues during lifetime.

Finally, the mC level observed in mouse cerebrum (cerebral cortex) remains roughly constant until p14. It then shows a much stronger increase with age compared to human cerebrum.

In summary, we provide the first data on age-dependent variations in global hmC levels in humans over the entire lifespan. After a strong increase during early postnatal stages and adolescence, hmC reaches a steady-state level of 1.2 % in the fully developed brain.^[8] This is nearly twice as high as in mouse cerebral cortex.^[4] Furthermore, global levels of fC in brain tissues of human adults, and cytosine modification levels in human cerebral occipital cortex neurons were quantified. We also report age-dependent variations in fC, hmC, and mC levels in mouse tissues between postnatal day 1 and postnatal month 18. Our measurements revealed that the level of fC decreases strongly at early developmental stages, showing an inverted age-dependent trend compared to hmC. Our data support the idea that fC in the developing brain is mainly linked to active DNA demethylation,^[2,3,5a,12d,e] while hmC in this context is rather a stable epigenetic mark.^[14] The observed trends in global cytosine-modification dynamics during the lifespan of an organism are conserved^[15] between mammalian species^[16] and are similar in appearance in different organs.^[4,6,9b,10a,b,e,11,16,17]

Keywords: 5-formylcytosine · 5-hydroxymethylcytosine · aging · epigenetics · cerebral cortex

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 12511–12514
Angew. Chem. **2015**, *127*, 12691–12695

- [1] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
- [2] a) T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150; b) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [3] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
- [4] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377; *Angew. Chem.* **2010**, *122*, 5503–5505.
- [5] a) D. Globisch, M. Münzel, M. Müller, S. Michalakakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367; b) A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181; c) T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljárova, D. Schürmann, S. Michalakakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574–581.
- [6] a) R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, *Science* **2013**, *341*, 1237905; b) C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68–72; c) K. E. Szulwach, X. Li, Y. Li, C.-X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, P. Jin, *Nat. Neurosci.* **2011**, *14*, 1607–1616; d) L. Chouliaras, D. L. A. van den Hove, G. Kenis, S. Keitel, P. R. Hof, J. van Os, H. W. M. Steinbusch, C. Schmitz, B. P. F. Rutten, *Curr. Alzheimer Res.* **2012**, *9*, 536–544; e) H. Chen, S. Dzitoyeva, H. Manev, *Restor. Neurol. Neurosci.* **2012**, *30*, 237–245; f) M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, *Cell Rep.* **2013**, *3*, 291–300.
- [7] a) W. Sun, L. Zang, Q. Shu, X. Li, *Genomics* **2014**, *104*, 347–351; b) M. Santiago, C. Antunes, M. Guedes, N. Sousa, C. J. Marques, *Genomics* **2014**, *104*, 334–340; c) L. Wen, F. Tang, *Genomics* **2014**, *104*, 341–346; d) D. L. A. van den Hove, L. Chouliaras, B. P. F. Rutten, *Curr. Alzheimer Res.* **2012**, *9*, 545–549.
- [8] T. F. J. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577–1590.
- [9] a) S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599; b) A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakakis, *Cell Rep.* **2015**, *11*, 283–294.
- [10] a) T. Wang, Q. Pan, L. Lin, K. E. Szulwach, C.-X. Song, C. He, H. Wu, S. T. Warren, P. Jin, R. Duan, X. Li, *Hum. Mol. Genet.* **2012**, *21*, 5500–5510; b) L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, X. Li, Y. Hou, H. Guo, X. Wu, C. He, R. Li, F. Tang, J. Qiao, *Genome Biol.* **2014**, *15*, R49; c) S.-G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **2011**, *39*, 5015–5024; d) N. Coppieters, B. V. Dieriks, C. Lill, R. L. Faull, M. A. Curtis, M. Dragunow, *Neurobiol. Aging* **2014**, *35*, 1334–1344; e) T. F. Kraus, V. Guibourt, H. A. Kretzschmar, *J. Neural Transm.* **2014**, DOI: 10.1007/s00702-00014-01346-00704; f) S. F. Field, D. Beraldi, M. Bachman, S. K. Stewart, S. Beck, S. Balasubramanian, *PLoS ONE* **2015**, *10*, e0118202.
- [11] a) K. D. Siegmund, C. M. Connor, M. Campan, T. I. Long, D. J. Weisenberger, D. Biniszkiwicz, R. Jaenisch, P. W. Laird, S. Akbarian, *PLoS ONE* **2007**, *2*, e895; b) D. G. Hernandez, M. A. Nalls, J. R. Gibbs, S. Arepalli, M. van der Brug, S. Chong, M. Moore, D. L. Longo, M. R. Cookson, B. J. Traynor, A. B. Singleton, *Hum. Mol. Genet.* **2011**, *20*, 1164–1172; c) L. Chouliaras, D. L. A. van den Hove, G. Kenis, S. Keitel, P. R. Hof, J. van Os, H. W. M. Steinbusch, C. Schmitz, B. P. F. Rutten, *Neurobiol. Aging* **2012**, *33*, 1672–1681.
- [12] a) S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620; b) S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1133; c) J. Guo, Y. Su, C. Zhong, G. Ming, H. Song, *Cell* **2011**, *145*, 423–434; d) A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338; e) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67–79.
- [13] S. Dzitoyeva, H. Chen, H. Manev, *Neurobiol. Aging* **2012**, *33*, 2881–2891.
- [14] a) M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049–1055; b) M. A. Hahn, P. E. Szabo, G. P. Pfeifer, *Genomics* **2014**, *104*, 314–323.
- [15] R. D. Almeida, V. Sottile, M. Loose, P. A. De Sousa, A. D. Johnson, A. Ruzov, *Epigenetics* **2012**, *7*, 137–140.
- [16] T. Zheng, Q. Lv, X. Lei, X. Yin, B. Zhang, *Neurochem. Res.* **2015**, *40*, 688–697.
- [17] S. A. Tammen, G. G. Dolnikowski, L. M. Ausman, Z. Liu, K.-c. Kim, S. Friso, S.-W. Choi, *J. Cancer Prev.* **2014**, *19*, 301–308.

Received: March 24, 2015

Revised: May 13, 2015

Published online: July 3, 2015

3.2 TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression

Arshan Perera³, David Eisen³, Mirko Wagner³, Silvia K. Laube, Andrea F. Künzel, Susanne Koch, Jessica Steinbacher, Elisabeth Schulze, Victoria Splith, Nana Mittermeier, Markus Müller, Martin Biel, Thomas Carell, Stylianos Michalaki. (³ geteilte Erstautorenschaft)

Prolog

Die Tet Enzyme (Tet 1-3) oxidieren 5-Methylcytosin (5mdC) zu 5-Hydroxymethyl- (5hmdC). In Neuronen korrelieren erhöhte 5hmdC Level mit erhöhter Genexpression. Die Mechanismen die dabei die Tet-Aktivität kontrollieren sind noch wenig verstanden. Bekannt ist, wie die neuronale Tet 3 Isoform, welcher die DNA-Bindedomäne fehlt, zur DNA gebracht wird. Um Tet 3 bindende Faktoren zu identifizieren suchten wir nach Proteinen die aus Maus Retina Lysaten koangereichert werden. Der Transkriptionsrepressor REST konnte als hoch angereicherter spezifischer Tet 3 Binder identifiziert werden. In einer Koexpression war REST dazu in der Lage die Hydroxylaseaktivität von Tet 3 zu erhöhen und Überexpression von Tet 3 führte zu einer Aktivierung der Transkription von REST Zielgenen. Darüber hinaus konnte gezeigt werden, dass Tet 3 zudem mit NSD3 und zwei weiteren H3K36 Methyltransferasen interagiert und die H3K36 Trimethylierung induziert.

Autorenbeitrag

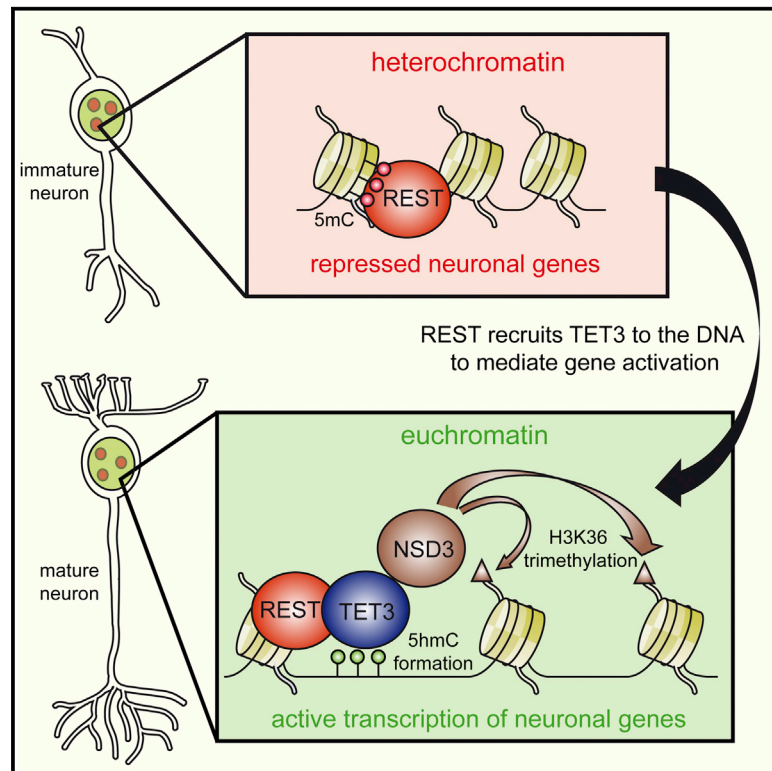
Der Verdau der DNA, die massenspektrometrischen Quantifizierung, sowie die Auswertung und Interpretation der Daten führte ich in enger Zusammenarbeit mit *Mirko Wagner* durch. Für eine weitere Auflistung des Beitrags der übrigen Autoren siehe Manuskript.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags.

TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression

Graphical Abstract



Authors

Arshan Perera, David Eisen, ..., Thomas Carell, Stylianos Michalakis

Correspondence

thomas.carell@cup.uni-muenchen.de (T.C.),
stylianos.michalakis@cup.uni-muenchen.de (S.M.)

In Brief

Neuronal differentiation involves major epigenetic changes, such as accumulation of 5-hydroxymethylcytosine (5hmC). Perera et al. found that TET3 interacts with transcriptional regulators and histone writers. They demonstrate that REST recruits TET3 for context-specific 5hmC formation and TET3 recruits NSD3 for H3K36 trimethylation.

Highlights

- TET3 interacts with transcriptional regulators and histone writers
- REST recruits TET3 and stimulates its hydroxylase activity
- TET3 activates context-specific gene transcription
- TET3 enhances the activity of NSD3 to generate H3K36 trimethylation

Accession Numbers

GSE65860
PXD001857



TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression

Arshan Perera,^{1,3} David Eisen,^{2,3} Mirko Wagner,^{2,3} Silvia K. Laube,² Andrea F. Künzel,² Susanne Koch,¹ Jessica Steinbacher,² Elisabeth Schulze,¹ Victoria Splith,¹ Nana Mittermeier,¹ Markus Müller,² Martin Biel,¹ Thomas Carell,^{2,*} and Stylianos Michalakis^{1,*}

¹Center for Integrated Protein Science Munich CiPSM at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

²Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

³Co-first author

*Correspondence: thomas.carell@cup.uni-muenchen.de (T.C.), stylianos.michalakis@cup.uni-muenchen.de (S.M.)

<http://dx.doi.org/10.1016/j.celrep.2015.03.020>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

Ten-eleven translocation hydroxylases (TET1-3) oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). In neurons, increased 5hmC levels within gene bodies correlate positively with gene expression. The mechanisms controlling TET activity and 5hmC levels are poorly understood. In particular, it is not known how the neuronal TET3 isoform lacking a DNA-binding domain is targeted to the DNA. To identify factors binding to TET3, we screened for proteins that co-precipitate with TET3 from mouse retina and identified the transcriptional repressor REST as a highly enriched TET3-specific interactor. REST was able to enhance TET3 hydroxylase activity after co-expression and overexpression of TET3-activated transcription of REST target genes. Moreover, we found that TET3 also interacts with NSD3 and two other H3K36 methyltransferases and is able to induce H3K36 trimethylation. We propose a mechanism for transcriptional activation in neurons that involves REST-guided targeting of TET3 to the DNA for directed 5hmC generation and NSD3-mediated H3K36 trimethylation.

INTRODUCTION

Epigenetic mechanisms are critically involved in neuronal differentiation and synaptic network formation. 5-hydroxymethylcytosine (5hmC), a recently identified epigenetic DNA modification, is present at high levels in the brain (Münzel et al., 2010) and accumulates during neuronal differentiation (Hahn et al., 2013). 5hmC is generated by ten-eleven translocation (TET) hydroxylases (TET1-3) from 5-methylcytosine (5mC) (Kriacounis and Heintz, 2009; Tahiliani et al., 2009). Accumulation of 5hmC in neuronal genes was shown to correlate positively with gene expression (Colquitt et al., 2013; Hahn et al., 2013; Mellén et al., 2012; Song et al., 2011; Szulwach et al.,

2011), suggesting that 5hmC plays a key role in neuronal differentiation (Santiago et al., 2014). How TET hydroxylase activity is controlled and directed to neuronal genes is not known. The TET1 isoform contains an amino-terminal CXXC domain that is thought to be important for binding to DNA (Xu et al., 2011). Tet2 lacks such a CXXC domain, but was shown to functionally interact with the CXXC domain containing proteins CXXC4 and CXXC5 (Ko et al., 2013; Williams et al., 2011).

TET3, the major TET isoform in neurons (Colquitt et al., 2013; Hahn et al., 2013), exists in various isoforms generated by alternative splicing (Liu et al., 2013), two isoforms containing a CXXC domain and one shorter isoform lacking such a domain. In particular, the TET3 lacking a CXXC domain (neuronal TET3) was shown to be enriched in neuronal tissue (Liu et al., 2013). It is not known how neuronal TET3 is targeted to the DNA. It was suggested that transcriptional regulators might bind directly to TET3, facilitating its specific targeting to genomic positions, and, hence, control 5hmC levels (Xu et al., 2012). Transcription-factor-dependent control of TET3 activity could potentially contribute to epigenetic regulation of gene expression in neurons. However, transcriptional regulators that functionally interact with TET3 have not been identified so far.

To directly address this issue, we investigated the function of TET3 and 5hmC during the formation of a complex neural network: the murine retina. We show that 5hmC levels increase during postnatal retinal network maturation. This gain of 5hmC results in an elevated expression of genes involved in important retinal and/or neuronal functions. To identify factors able to regulate TET3 activity and 5hmC formation in a gene-specific manner, we purified retinal nuclear proteins interacting with TET3. We found that the transcriptional regulator REST specifically interacts with the neuronal TET3 isoform and elevates its hydroxylase activity. Moreover, we show that TET3 overexpression induces the expression of REST target genes. We postulate that transcriptional activation of specific genes results from de novo chromatin remodeling that involves activity of histone 3 lysine 36 methyltransferases like NSD3.

RESULTS

5hmC Is Acquired during Postnatal Retinal Development

Development of neural networks is a complex process that involves the generation, positioning, and synaptic wiring of neurons. This developmental process is finalized with activity-dependent maturation of neurons and refinement of their network. In the mammalian retina, this terminal maturation process involves major morphological and functional changes, largely occurring within a week after eye opening (Hoon et al., 2014; Okawa et al., 2014). These morphological and functional refinements go along with significant changes in the gene expression pattern. To visualize these changes, we applied label-free quantification (LFQ) comparing the retinal protein levels in 2-week-old mice (eye opening) with 3-week-old mice (mature state; Figure 1A), and we found a significant upregulation of proteins involved in synaptic function and visual processing over time (Figures 1A and S1A; Table S1). The molecular mechanisms controlling the underlying transcriptional changes are only partially understood (Gregory-Evans et al., 2013; Xiang, 2013).

We hypothesized that epigenetic regulation at the level of DNA might be involved. First evidence comes from the finding that the levels of the epigenetic DNA modification 5hmC in brain neurons increase with age (Hahn et al., 2013; Münzel et al., 2010). To test for its involvement in terminal retinal maturation, we analyzed the levels of 5hmC during this period. We first used a 5hmC-specific antibody for immunolocalization studies on retinal cryosections and observed a strong increase of the 5hmC immunosignal in all nuclear layers of the retina between week 2 and week 3 (Figures 1B and 1C). At eye opening (week 2), we observed high levels of 5hmC in nuclei of cells within the inner nuclear layer (INL) and the ganglion cell layer (GCL). In contrast, photoreceptor nuclei located in the outer nuclear layer (ONL) were immunonegative (Figure 1B). Strikingly, only 1 week after eye opening (week 3), the levels of 5hmC increased dramatically in both the INL and the GCL. Moreover, rod and cone photoreceptors in the ONL were now positive for 5hmC (Figures 1C and 1D). Confirming the specificity of the anti-5hmC antibody, the immunosignal was quantitatively depleted by 2.5 μ M 5hmC-containing DNA oligonucleotides (Figure 1C, inset). At the sub-nuclear level, most of the 5hmC signal was found to localize in euchromatic regions in both conventional and inverted nuclei (Figures 1E and 1F). In contrast, we found 5mC to be concentrated in pericentromeric heterochromatin (Figures 1E and 1F), suggesting that while 5mC is present in transcriptionally silenced regions, 5hmC is associated with transcriptional activity.

To quantify the levels of 5hmC along with all other cytosine modifications in mouse genomic DNA, we combined ultra-high-performance liquid chromatography with tandem mass spectrometry (UHPLC-MS/MS) (Schiesser et al., 2013; Table S7). The levels of 5hmC increased significantly (Figure 1G and S1B) ($p < 0.005$, Student's *t* test), whereas the levels of the further oxidized derivative 5-formylcytosine (5fC) dropped significantly during the same time frame (Figures 1I and S1C), suggesting distinct roles for 5hmC and 5fC in neurons. Interestingly, the levels of 5mC remained stable during this developmental window (Figures 1H and S1D).

Thus, retinal cells specifically accumulate genomic 5hmC levels during terminal retinal maturation. Importantly, retinal cells

at this developmental stage do not divide; thus, this effect cannot result from increased DNA replication.

To investigate the dynamics of 5hmC and 5mC formation in non-dividing cells, we performed an isotope-tracing experiment by culturing retinal explants from 11-day-old mice for 12 days in the presence of the labeled methyl group donor [$^{13}\text{CD}_3$]-methionine, which provides labeled 5mC (5-[$^{13}\text{CD}_3$]-mC) and 5hmC (5-[$^{13}\text{CD}_2$]-hmC) that can be quantified using UHPLC-MS/MS (Table S8). Confirming the presence of de novo methylation in terminally differentiated retinal cells, we detected 5-[$^{13}\text{CD}_3$]-mC at linearly increasing levels (slope: 0.531, $R^2 = 0.999$) reaching a value of 3.9% of total 5mC after 9 days of [$^{13}\text{CD}_3$]-methionine feeding (Figure 1J). We further detected [$^{13}\text{CD}_2$]-labeled 5hmC (Figure 1K). However, the level of 5-[$^{13}\text{CD}_2$]-hmC was low with only up to 1.6% after 9 days (slope: 0.208, $R^2 = 0.997$). Moreover, the majority of de novo 5hmC was generated from pre-existing and hence unlabeled 5mC (Figure S1E).

5hmC Accumulates in Neuronal Genes and Leads to Increased Gene Expression

To assess the genomic distribution of 5hmC during retinal maturation, we performed antibody-based 5hmC DNA immunoprecipitation (hMeDIP) from 2- and 3-week-old retina followed by next-generation sequencing (NGS). In agreement with the immunohistochemical and UHPLC-MS/MS quantification data, we identified 5,921 intragenic regions that gained 5hmC signal and approximately 20 times less with lost 5hmC signal (Figures 2A and S2A; Table S9). Gene ontology (GO) analysis revealed that genes involved in the morphology of nervous system, CNS development, differentiation of neurons, neuritogenesis, or guidance of axons are significantly enriched within the group of genes that gained 5hmC after eye opening (from week 2 to week 3) (Figure 2B). The genes enriched in the hMeDIP analysis were positively correlated to those enriched in the group of proteins with elevated expression in the mature retina ($p = 6.2 \times 10^{-11}$, Fischer's exact test; Figures 1A and S2B; Table S1). Moreover, 5hmC marks were particularly enriched in gene bodies (Figure 2C), suggesting a regulatory function of 5hmC in gene expression. To investigate if gain of 5hmC is associated with active transcription, we correlated the hMeDIP data with the LFQ protein expression data and found that gain of 5hmC levels from week 2 to week 3 resulted in significantly increased protein expression of the corresponding genes (Figure 2D).

Neuronal TET3 Interacts with Transcriptional Regulators and Histone Writers

5hmC is generated from 5mC by TET hydroxylases. To study if the elevated 5hmC levels in the mature retina result from increased TET-mediated oxidation, we analyzed the expression profile of the three TET enzymes during postnatal development of the mouse retina. We first found that TET3 is the dominant isoform in the retina and its levels increase at eye opening (Figure 3A; Table S10). Importantly, the expression levels at eye opening (postnatal day 13, p13) and at the mature stage (p20) were comparable, ruling out that increased Tet expression causes the increased 5hmC levels (Figure 3A).

TET3 exists in three isoforms that differ in their N-terminal sequence: two isoforms with a CXXC DNA-binding domain and

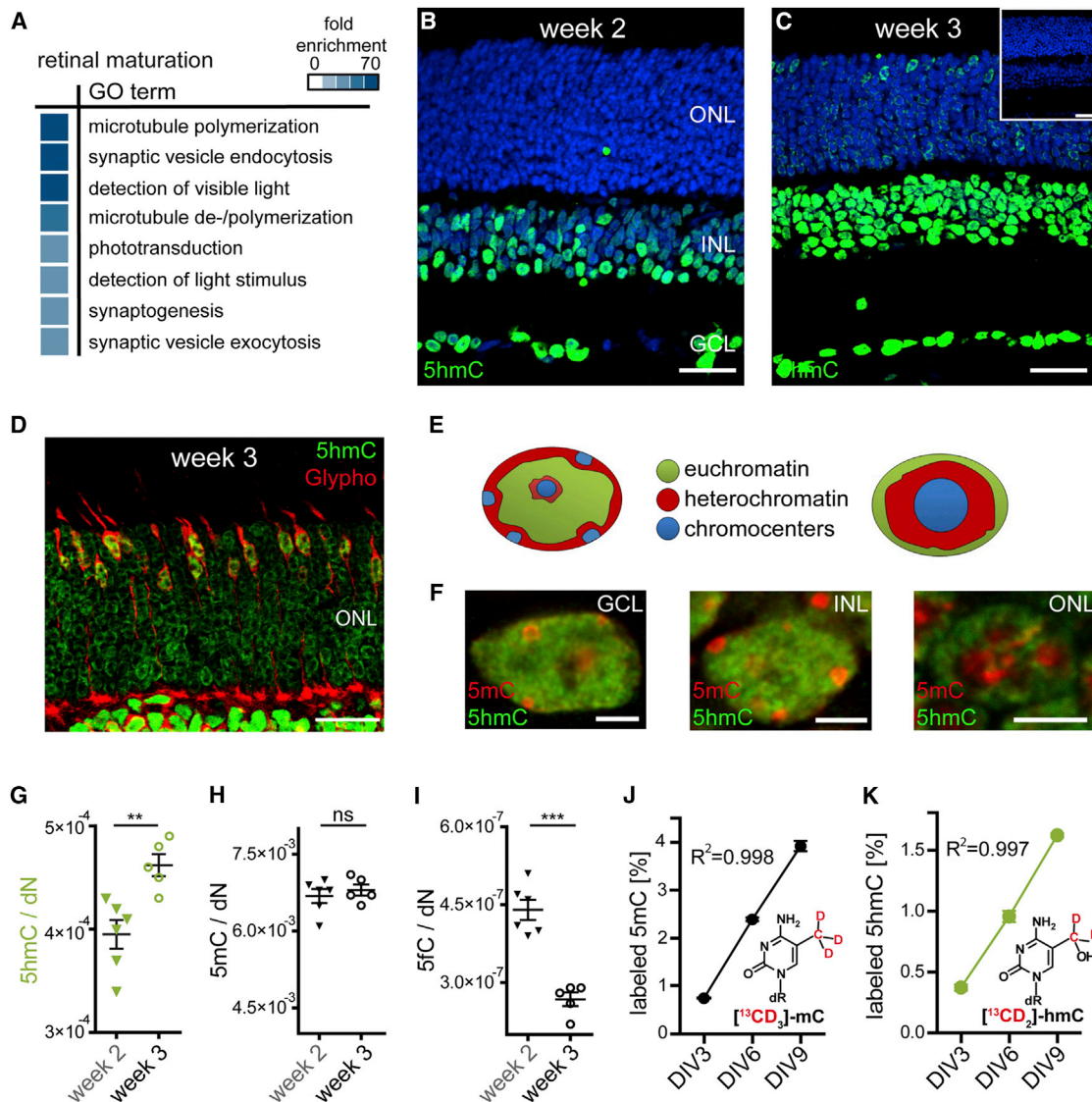


Figure 1. Transcriptional and Epigenetic Changes during Retinal Maturation

(A) GO analysis of upregulated proteins during retinal maturation (from eye opening at week 2 to terminal maturation of the retina at week 3) identified by LFQ is shown.

(B and C) Confocal scans from mouse retinal slices at week 2 (B) and week 3 (C) immunolabeled for 5hmC (green). There is a marked increase in 5hmC signal in nuclei of all retinal layers from week 2 to week 3, in particular in photoreceptors (ONL). The inset in (C) is from a depletion experiment proving the specificity of the anti-5hmC antibody.

(D) Magnification of ONL stained for 5hmC (green) and glycogen phosphorylase (Glypho) (red) is shown.

(E) Schematic view of conventional and inverted nuclei depicts heterochromatin and euchromatin regions.

(F) High-magnification images of conventional nuclei within the GCL or the INL and a rod nucleus with inverted nuclear architecture (Solovei et al., 2009) co-immunolabeled with 5hmC (green) and 5mC (red). In all cases, 5hmC is mainly found in euchromatin whereas 5mC is localized in heterochromatic regions.

(G–I) Scatter plots of UHPLC-MS/MS quantification of global (G) 5hmC, (H) 5mC, and (I) 5fC in the mouse retina at weeks 2 and 3 reveal an age-dependent increase in 5hmC levels, stable 5mC levels, and decreasing 5fC levels.

(J and K) Isotope tracing combined with UHPLC-MS/MS in mouse retinal explant cultures. Retinal explant cultures were fed with [methyl- $^{13}\text{CD}_3$]-L-methionine and harvested for UHPLC-MS/MS quantification of labeled 5mC (J) and 5hmC (K) after 3, 6, and 9 days in vitro (DIV). Summary data are mean \pm SEM in (G–I) and mean \pm SD in (J and K). ** $p < 0.01$, *** $p < 0.001$ (Student's t test). The cell nuclei in (A and B) were stained with Hoechst 3442 nuclear dye (blue). Scale bar, 25 μ m in (B–D) and 3 μ m in (F). See also Figure S1 and Tables S7 and S8.

a short isoform enriched in neuronal cells lacking the CXXC domain (Liu et al., 2013). Using isoform-specific primer sets, we found that the short variant lacking the CXXC domain is the

major TET3 isoform in the retina (Figure 3B). Other than the isoforms with the CXXC domain, this neuronal TET3 variant has no DNA-binding affinity (Xu et al., 2012). We hypothesized that

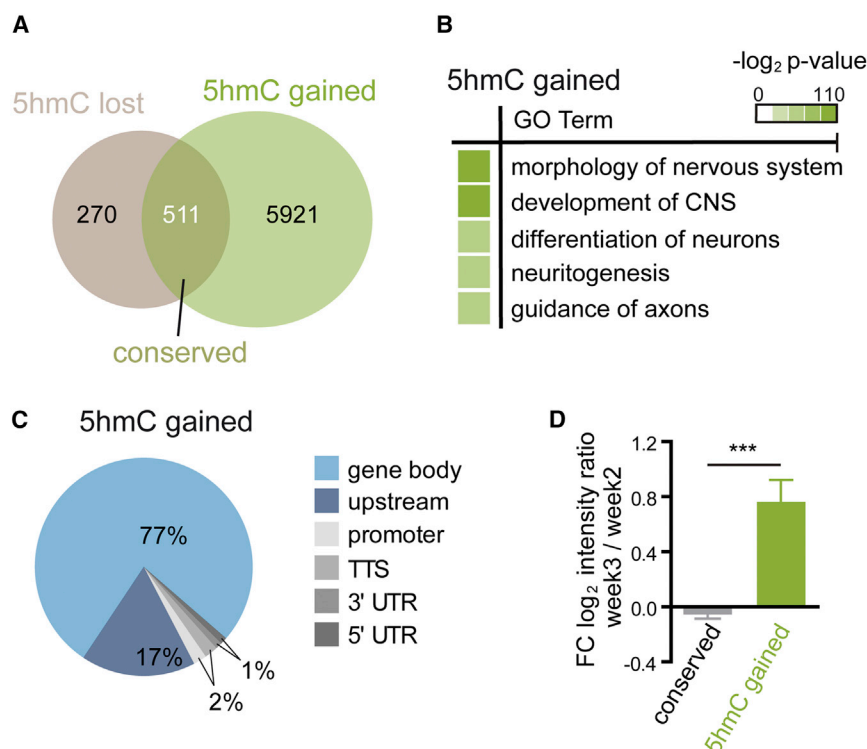


Figure 2. Genomic Localization of 5hmC in the Mouse Retina and Its Effect on Gene Expression

(A) Venn diagram shows 5hmC dynamics during retinal maturation (postnatal week 3/postnatal week 2). (B) GO analysis of genes that gained 5hmC during retinal maturation is shown. (C) Pie chart shows intragenic distribution of 5hmC regions in mature retina. (D) LFQ of proteins encoded by genes with conserved (gray) or gained (green) 5hmC levels during retinal maturation. Summary data are mean \pm SEM. ***p < 0.001 (Student's t test). TSS, transcription starting site; TTS, transcription termination site. See also Figure S2 and Table S9.

unknown DNA-binding (transcription) factors could bind and target neuronal TET3 to pre-existing 5mC for subsequent oxidation to 5hmC (Figure 3C). To identify nuclear proteins that have the potential to specifically bind to neuronal TET3 in mouse retina, we combined affinity purification with mass spectrometry (MS). To this end, we transduced mouse retinal explant cultures with lentiviral vectors expressing an eGFP-fusion protein of the neuronal TET3 splice variant. Control cultures were transduced with lentiviral vectors expressing eGFP only. After onset of expression, nuclear protein complexes were extracted, affinity-purified, and subsequently labeled with TMT 2-plex reagents in forward and reverse label-swap experiments for comparative quantification using LC-MS (see the Experimental Procedures). These experiments allowed us to identify 52 proteins that were significantly enriched for binding to TET3 (>1.5-fold enrichment) in forward and reverse labeling experiments (Figure 3D, bottom right quadrant). Notably, nine of those proteins, such as REST, SUZ12, and CEP290, could be classified as regulators of transcription (Figure 3D; Table S2). Another six proteins are histone variants involved in chromatin modification, i.e., H3F3A and H2AFJ (Figure 3D; Table S2).

TET3 is endogenously expressed in the mouse retina (Figure 3A). Thus, TET3-eGFP is expected to compete for binding partners with endogenous TET3. To identify proteins that bind to endogenous TET3, we repeated the affinity purification experiment using a TET3-specific antibody. Endogenous TET3 was highly enriched by the antibody highlighting its specificity (Figure 3E). Moreover, we identified five transcriptional regulators (ASXL1, CTCF, MORF4L1, REST, and VAX1), three H3K36 methyltransferases (NSD2, NSD3, and SETD2), and histone 3 (H3F3A)

as interactors with significant enrichment for binding to endogenous TET3 (>1.5-fold) (Figure 3E; Table S3).

Interestingly, the TET3-interacting protein with the highest enrichment score was the transcriptional repressor REST (Figure 3E), which was also significantly enriched in the TET3-eGFP affinity purification experiment (Figure 3D). To verify the TET3-REST interaction, we performed a reverse experiment using a REST-specific antibody that efficiently immunoprecipitates endogenous REST from mouse retina (Figure 3F; Table S4). Importantly, in this experiment, the REST antibody affinity-purified TET3 very efficiently with the overall third-highest enrichment score, but failed to immunoprecipitate TET1 and TET2 (Figure 3F). Moreover, none of the identified TET3 peptides was specific for the long CXXC-containing isoforms, suggesting that REST interacts specifically with the neuronal TET3 isoform. Taken together, these experiments identify REST as a major specific interacting protein of neuronal TET3.

REST Recruits Neuronal TET3 to Mediate 5hmC Formation and Transcriptional Activation

To determine the functional effects of REST on TET3 hydroxylase activity, we overexpressed TET3-eGFP alone and together with REST in HEK293T cells and subsequently quantified the 5hmC levels (Figure 4A). Non-transfected HEK293T cells have very low endogenous 5hmC levels, and overexpression of TET3-eGFP alone resulted in increased 5hmC levels (Figure 4A). Importantly, however, after co-expression with REST, TET3-eGFP generated significantly higher levels of 5hmC (Figure 4A). Conversely, after small interfering RNA (siRNA)-mediated knock-down of endogenous REST, overexpression of TET3-eGFP led to a significantly weaker elevation of 5hmC levels compared to control siRNA treatment (Figure 4B). Thus, REST is both necessary and sufficient to elevate TET3 hydroxylase activity.

REST is a transcriptional repressor that binds to specific RE-1 binding sites present in regulatory regions of its target genes (Kraemer et al., 1992). Binding of REST results in transcriptional repression of the corresponding genes (Chong et al., 1995; Schoenherr

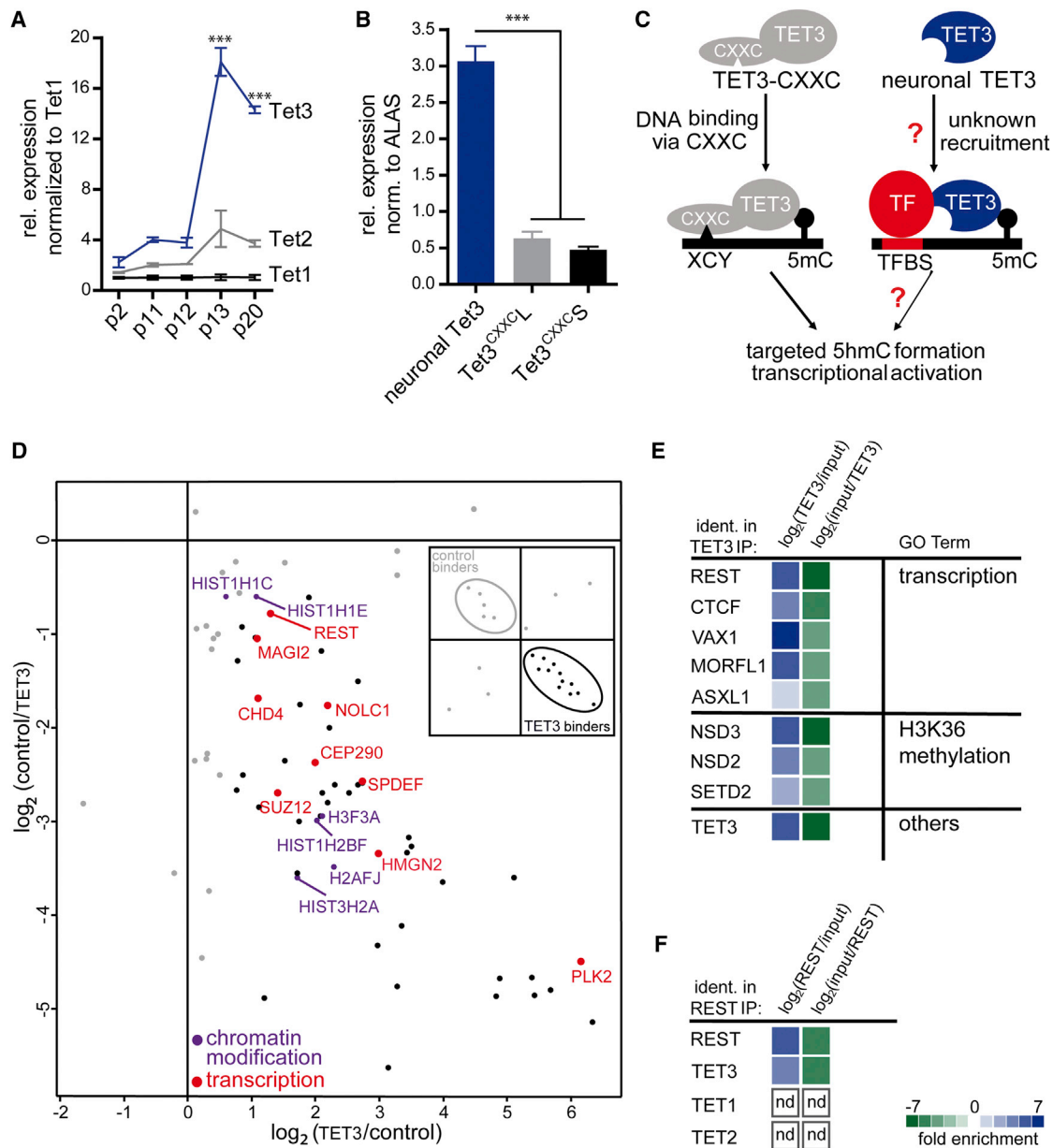


Figure 3. Tet Expression Levels and Identification of TET3-Interacting Proteins in Retina

(A) qRT-PCR analysis of Tet1, Tet2, and Tet3 gene expression in mouse retina at postnatal day (p) 2, 11, 12, 13, and 20 normalized to corresponding Tet1 levels. Tet3 is expressed significantly higher than Tet1 and Tet2 at p13 and p20.

(B) qRT-PCR analysis of Tet3 isoforms. Tet3 lacking a CXXC DNA-binding domain (neuronal Tet3) is the main isoform in retina.

(C) Cartoon illustrating putative DNA-binding mechanisms of TET3 isoforms lacking (neuronal TET3) or containing a CXXC domain. We hypothesized that neuronal TET3 is recruited to the DNA by transcription factors for context-specific hydroxymethylation and induction of gene expression.

(D) Scatter plot is shown of TET3 interactors identified using lentiviral TET3-eGFP overexpression (OE) in retinal explant cultures followed by GFP-trap and LC-MS/MS.

(E) Heatmap of endogenous TET3 interaction partners identified by immunoprecipitation and LC-MS/MS is shown.

(F) Heatmap of endogenous REST immunoprecipitation LC-MS/MS experiment confirms the interaction with endogenous TET3 (nd, not detected). Summary data are mean \pm SEM in (A and B). *** $p < 0.001$ (one-way ANOVA). See also Table S10.

and Anderson, 1995). We therefore asked whether TET3 might affect gene expression of REST target genes. To address this question, we overexpressed TET3-eGFP in HEK293T cells that endogenously contained REST (Dietrich et al., 2012), and we

determined the global protein levels using LFQ (Figures 4C–4E). Overexpression of TET3-eGFP resulted in significant upregulation of 40 proteins, and 26 of those were previously non-expressed (proteins with infinite log2 ratios in Figure 4E).

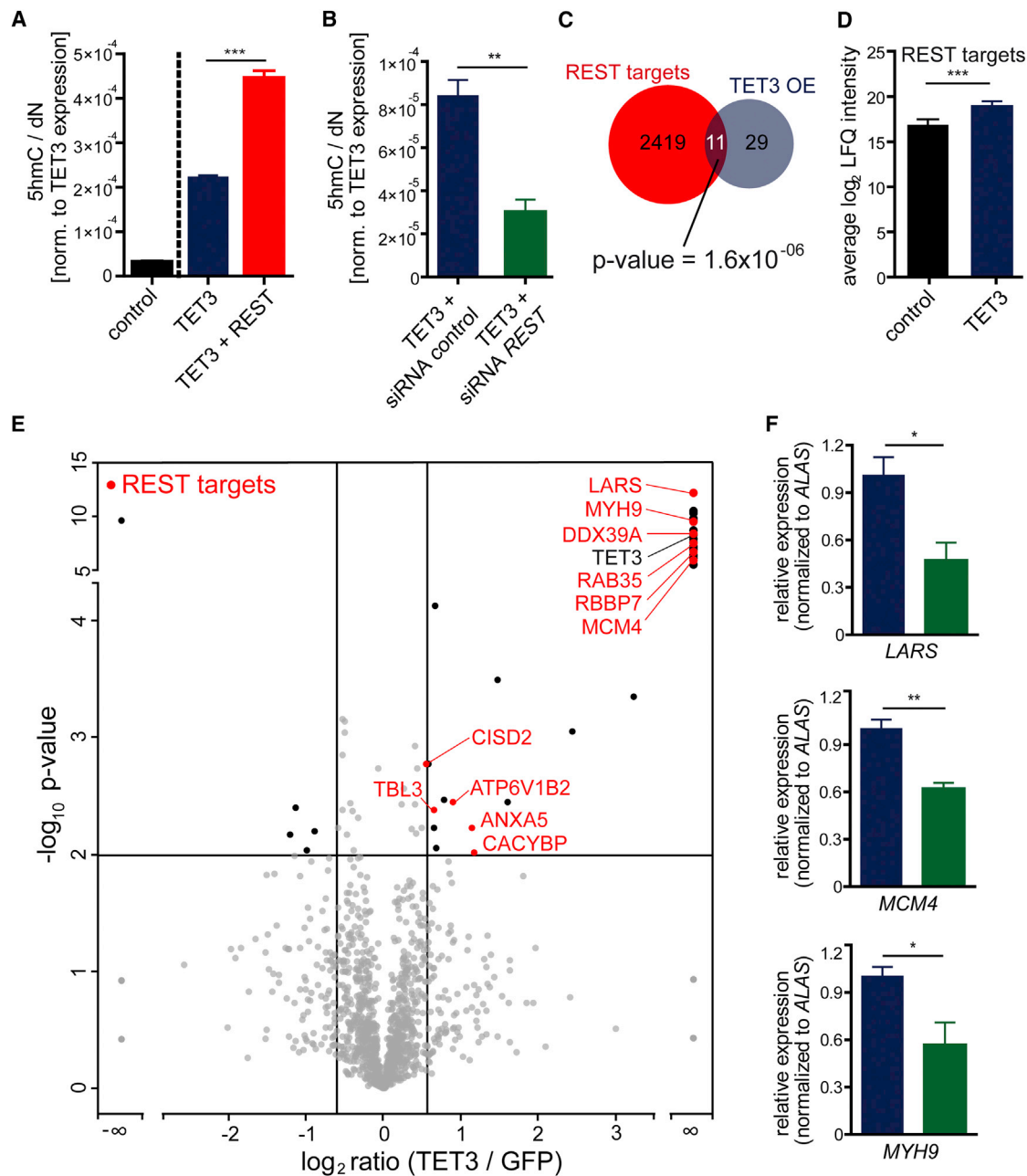


Figure 4. Functional Analysis of the Interaction of Transcriptional Repressor REST with Neuronal TET3

(A) UHPLC-MS/MS quantification of global 5hmC levels after OE of TET3-eGFP or REST and TET3-eGFP is shown.

(B) UHPLC-MS/MS quantification of global 5hmC levels after OE of TET3-eGFP in the presence of siRNA directed against human REST or a scrambled siRNA control (ctrl) is shown.

(C) Identification of global REST target gene expression using LFQ after eGFP (control) or TET3-eGFP OE in HEK293T cells. REST target genes are expressed significantly higher after TET3 OE.

(D) Venn diagram shows the significant overlap (Fisher's exact test; p value = 1.6×10^{-6}) of proteins enriched after TET3 OE and REST target genes in HEK293T cells.

(E) Volcano plot of protein expression ratios between TET3-eGFP and eGFP OE experiments in HEK293T cells as a function of statistical significance (Student's t test p value ≤ 0.01). Proteins with no statistically significant difference in expression between subsets are gray. Proteins with no detectable signal in one of the subsets were assigned a ratio of infinity. REST target genes are highlighted in red.

(F) Quantification of gene expression of three TET3-induced REST target genes in HEK293T cells using qRT-PCR after OE of TET3-eGFP in the presence of siRNA directed against human REST (green) or a scrambled siRNA control (ctrl, blue). Summary data in (A–C and F) are mean \pm SEM. *p < 0.05, **p < 0.005, ***p < 0.001 (Student's t test). See also Table S11.

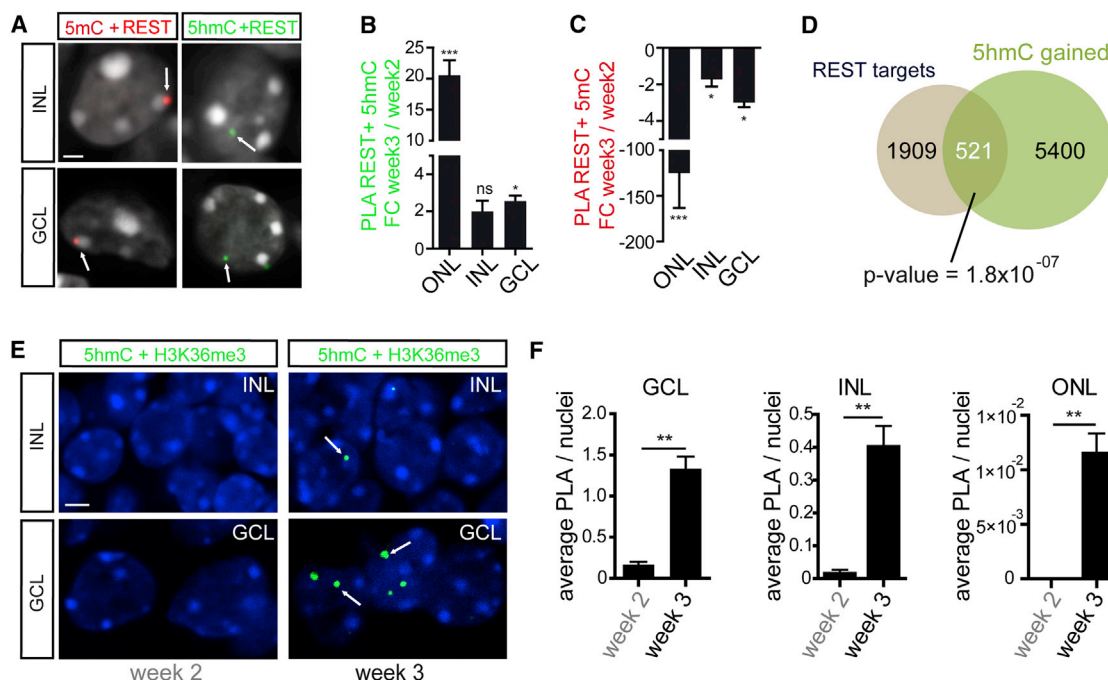


Figure 5. Accumulation of 5hmC in REST Target Genes Is Associated with Chromatin Remodeling

(A) PLA signal for 5mC and REST (red) is localized in heterochromatin. In contrast, the PLA signal for 5hmC and REST is localized in euchromatin (green).

(B and C) Quantification data of the 5hmC/REST (B) and 5mC/REST (C) PLA experiments in (A) are shown.

(D) Venn diagram shows the significant overlap (Fisher's exact test; p value = 1.8×10^{-07}) of REST target genes and 5hmC-gained genes during retinal maturation.

(E) PLA signal for 5hmC and H3K36me3 (green) in nuclei of the INL (top) and the GCL (bottom) at week 2 (left) and week 3 (right).

(F) Quantification data of the 5hmC/H3K36me3 PLA experiments in (E) are shown. Summary data in (B, C, and F) are mean \pm SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (Student's t test). Scale bar, 1 μ m in (A) and 5 μ m in (E).

Interestingly, 308 of the identified proteins were known REST targets (Lu et al., 2014), and the average log₂ LFQ intensity of those REST targets was significantly higher after TET3 overexpression compared to the eGFP control (Figure 4C). Importantly, 11 of the 40 TET3-induced proteins ($p = 1.6 \times 10^{-6}$, Fischer's exact test) were REST targets (Figure 4D; Table S5) and six of those (DDX39A, LARS, MCM4, MYH9, RAB35, and RBBP7) were previously non-expressed proteins (Figure 4E; Table S5).

To test if this TET3-mediated induction of REST target genes in HEK293T cells depends on REST, we overexpressed TET3-eGFP in the presence of REST-specific siRNA or scrambled control siRNA and measured the transcript levels of *LARS*, *MCM4*, and *MYH9* (Table S11). As shown in Figure 4E, all three REST target genes were exclusively expressed in the presence of TET3. Interestingly, in all three cases, the positive effect of TET3 on gene expression was significantly lower after knock-down of endogenous REST (Figure 4F), confirming that TET3 requires the presence of REST to de-repress REST target genes.

Accumulation of 5hmC in REST Target Genes Is Associated with Chromatin Remodeling

These results indicate that REST directs TET3 to repressed target genes and subsequent hydroxylation of pre-existing 5mC, then leads to de-repression and transcriptional activation of the genes. Repressed genes are usually found in hetero-

chromatin, whereas actively transcribed genes are located in euchromatin. As shown in Figure 1F, 5mC is enriched in heterochromatic structures and 5hmC in euchromatin. To test for the co-localization of REST with 5mC and 5hmC in nuclei of retinal neurons, we applied an in situ proximity ligation assay (PLA) combining REST-specific with 5mC- or 5hmC-specific antibodies. PLA gives only a positive signal when 5mC or 5hmC and REST are in close proximity (<40 nm) (Söderberg et al., 2006). We observed a positive PLA signal for both combinations, REST/5mC as well as REST/5hmC (Figure 5A). Importantly, the PLA signal for REST/5mC was localized in heterochromatin (Figure 5A, top), whereas the REST/5hmC signal was observed in euchromatin (Figure 5A, bottom). Thus REST-binding sites show distinct subnuclear localization depending on the oxidation status of adjacent genomic cytosines. Interestingly, the levels of PLA signal for REST/5mC and REST/5hmC varied significantly between week 2 and week 3 (Figures 5B and 5C). In particular, the REST/5hmC PLA signal correlated positively with the 5hmC levels increasing from week 2 to week 3 (Figure 5B). In contrast, the PLA signal for REST/5mC showed an inverse behavior, decreasing during the same period of time (Figure 5C). Together, these results suggest a functional correlation between REST-binding sites and proximal epigenetic DNA marks that might affect chromatin state and gene expression.

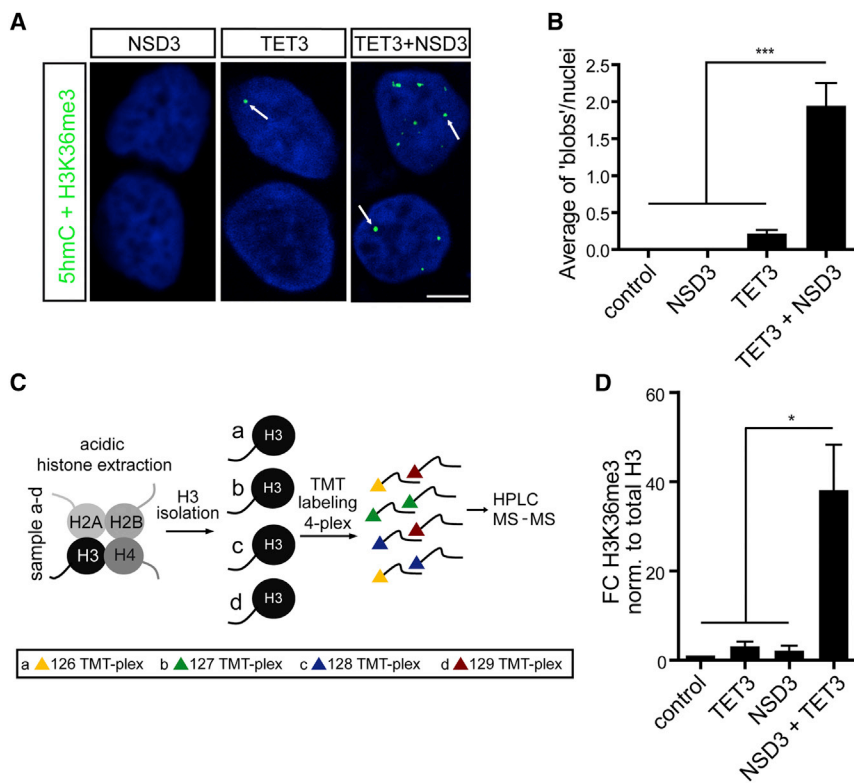


Figure 6. Effect of TET3-Mediated Hydroxylation on Chromatin Structure

(A) PLA signal for 5hmC and H3K36me3 in HEK293T cells overexpressing NSD3 (left), TET3-eGFP (middle) and TET3 + NSD3 (right). (B) Quantification data of the 5hmC/H3K36me3 PLA experiments in (A) are shown. (C) Scheme of TMT⁴-plex isobaric labeling and LC-MS quantification of H3 modifications is shown. (D) Isobaric labeling LC-MS quantification of H3K36me3 in HEK293T cells overexpressing TET3, NSD3, or TET3 and NSD3 compared to non-transfected cells. Summary data are mean \pm SEM in (B and D). * p < 0.05, *** p < 0.001 (one-way ANOVA). Scale bar, 5 μ m in (A).

tone H3 at lysine 36 (H3K36), which is an epigenetic mark commonly associated with transcription of active chromatin (Bernstein et al., 2005; Heintzman et al., 2007; Kouzarides, 2007).

TET3 Regulates NSD3 H3K36 Trimethylation Activity

To test if the TET3/NSD3 interaction is functionally relevant, we performed a PLA experiment with HEK293T cells transfected with TET3 and/or NSD3,

focusing on the co-localization of the products of the two enzymes, 5hmC and H3K36me3, respectively. NSD3 overexpression did not result in any detectable PLA signal (Figures 6A and 6B), which is in line with the very low levels of 5hmC in HEK293T cells (see Figure 4A). TET3-overexpressing cells showed only a few PLA signals, indicating that 5hmC elevation alone does not automatically result in 5hmC/H3K36me3 co-localization (Figures 6A and 6B). However, when both TET3 and NSD3 were overexpressed, we observed a synergism that resulted in a significant increase of the 5hmC/H3K36me3 PLA signal (Figures 6A and 6B). We next isolated and enriched histone 3 and quantified the extent of K36 trimethylation of histone 3 using isobaric labeling and MS (Figure 6C). This experiment showed a 40-fold increase in the levels of H3K36me3 in the HEK293T cells co-overexpressing TET3 and NSD3 compared to non-transfected cells (Figure 6D). Overexpression of TET3 or NSD3 alone provided only about a 2-fold increase of the H3K36me3 activation mark (Figure 6D). Thus, TET3 has a synergistic effect on the lysine methyltransferase activity of NSD3.

TET3 Overexpression in Retina Leads to Neuronal Gene Activation and H3K36me3

To analyze if TET3 overexpression has an effect on global gene expression in the retina, we transduced retinal explant cultures with lentiviral vectors expressing TET3-eGFP. Compared to retinal explant cultures overexpressing eGFP only, we identified 981 proteins that were upregulated after TET3-eGFP overexpression (Table S6). The TET3-eGFP overexpression strongly affects gene expression in the retina, resulting in the upregulation of proteins involved in several neurological functions (Figure 7A).

To test if REST has an effect on the genomic 5hmC content in the retina, we re-analyzed our 5hmC mapping data (see Figures 2A–2C) by searching for known REST target genes (Lu et al., 2014) within the subset of genes that gained 5hmC marks from week 2 to week 3 (Figure 5D). In support of a positive effect of REST on genomic 5hmC levels in the retina, we found a significant enrichment of REST target genes within the group of genes that gained 5hmC during retinal maturation ($p = 1.8 \times 10^{-7}$, Fischer's exact test) (Figure 5D).

Transcriptional activation of previously silenced genes requires chromatin remodeling, e.g., a switch from a heterochromatic to a euchromatic state (Voss and Hager, 2014). It was suggested that REST-mediated chromatin remodeling involves changes of the activating histone mark lysine 36 trimethylation of the nucleosomal histone H3 (H3K36me3) (Zheng et al., 2009). To test for this chromatin remodeling during retinal maturation, we analyzed the co-localization of 5hmC with H3K36me3 using specific antibodies in combination with PLA. As shown in Figures 5E and 5F, we observed a clear increase of the PLA signal for 5hmC/H3K36me3 co-localization in nuclei of 3-week-old retina, whereas at 2 weeks only a few or no PLA signals were observed. This increase in 5hmC/H3K36me3 PLA signal was observed in nuclei of all three major retinal layers (Figure 5F). Thus, from week 2 to week 3, the chromatin structure in retinal nuclei changed in a way that led to co-localization of 5hmC with the active chromatin mark H3K36me3.

Notably, three of the identified TET3-interacting proteins were the post-SET domain containing histone-lysine N-methyltransferases NSD3, NSD2, and SETD2 (Figure 3E). These histone writers are involved in the (tri-) methylation of nucleosomal his-

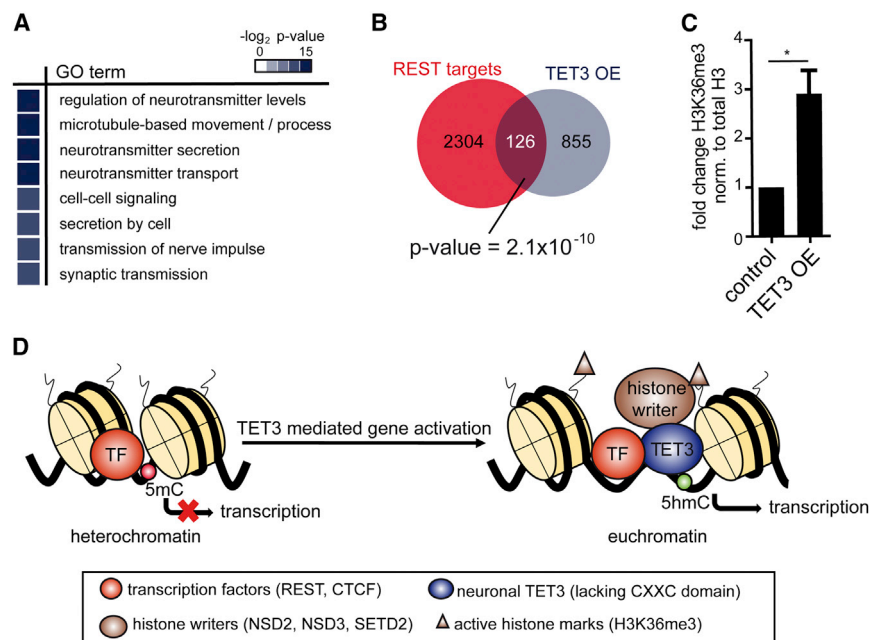


Figure 7. TET3 Overexpression Leads to Activation of Neuronal Genes and H3K36 Trimethylation during Retinal Maturation

(A) GO analysis reveals strong enrichment of neuronal proteins within the subset of proteins induced after TET3 OE in mouse retinal explants. (B) Venn diagram shows the significant overlap (Fisher's exact test; p value = 2.1×10^{-10}) of proteins enriched after TET3 OE and REST target genes.

(C) Isobaric labeling LC-MS quantification of H3K36me3 in mouse retina after TET3-eGFP or eGFP (control) OE. Summary data are mean \pm SEM. *p < 0.05 (Student's t test).

(D) Model for neuronal TET3-mediated transcriptional activation and chromatin remodeling. TET3 is recruited to the DNA by transcriptional regulators like REST for context-specific 5mC hydroxylation. Subsequently, TET3 mediates H3K36 trimethylation by recruitment of histone writers such as NSD3.

Correlation analysis again revealed that 126 known REST targets were significantly enriched in the subset of 981 TET3-upregulated proteins ($p = 2.1 \times 10^{-10}$, Fischer's exact test) (Figure 7B; Table S6).

To test whether TET3 also has a positive effect on lysine methyltransferase activity in the retina, we transduced retinal explant cultures with TET3-eGFP or eGFP using lentiviral vectors and determined the levels of H3K36me3 using isobaric labeling and MS. Overexpression of TET3-eGFP was able to elevate the levels of H3K36me3 by a factor of 3.8, suggesting that TET3-mediated hydroxylation of 5mC in retinal cells results in activation of NSD3 or similar enzymes to induce H3K36 trimethylation (Figure 7C). Our data indicate that this involves, at least in part, REST-assisted, TET3-mediated oxidation of genomic 5mC and subsequent NSD3-mediated chromatin remodeling.

DISCUSSION

Our comprehensive analysis of proteins interacting with the 5hmC-generating enzyme TET3 revealed a mechanism for specific epigenetic regulation of gene expression. We show that the transcription factor REST, a master regulator of genes involved in neuronal maturation processes (McGann et al., 2014), binds and recruits TET3 to the DNA for specific gene activation. In particular, we find that (1) REST regulates TET3 hydroxylase activity, (2) REST target genes accumulate 5hmC during retinal maturation, and (3) overexpression of TET3 activates REST target genes. In support of these findings, a recent study showed that deletion of REST in mouse embryonic stem cells (ESCs) led to significant loss of genome-wide 5hmC levels in regulatory regions (Feldmann et al., 2013). REST did not interact with TET1 or TET2. In addition, none of the TET3 peptides identified by MS after affinity purification of REST complexes from

mouse retinal lysates were specific for the TET3 isoforms with CXXC domain. Thus, REST-mediated recruitment of

TET3 seems to be specific for the neuronal TET3 isoform lacking a CXXC domain.

REST originally was described as a transcriptional repressor of neuronal genes in non-neuronal cells (Chong et al., 1995; Schoenherr and Anderson, 1995). In addition, REST was shown to be a key repressor in retinal ganglion cell commitment by preventing premature expression in retinal progenitor cells (Mao et al., 2011). It was hypothesized that, in neural progenitors, REST's repressive function is linked to gene regulatory networks related to neuronal differentiation (Mao et al., 2011). To act as a repressor, REST interacts with other proteins like mSin3a or COREST1 to form REST repressor complexes (Huang et al., 1999; McGann et al., 2014; Naruse et al., 1999). Importantly, none of the known REST-repressive complex proteins were identified in our REST-specific affinity purifications from mouse retinal lysates. There is now growing evidence suggesting that REST also plays a role in proper function of neuronal cells (Lu et al., 2014; Lunyak and Rosenfeld, 2005). For example, it was shown that REST functions as an activator of neuronal genes during neuronal maturation in the presence of a small double-stranded non-coding RNA corresponding to the NRSE/RE1 REST-binding sequence (Kuwabara et al., 2004).

Based on our 5hmC-mapping data, approximately 9% (521 of 5,921) of the genes that gained 5hmC during retinal maturation were known REST target genes. Thus, although REST contributes significantly ($p = 1.8 \times 10^{-7}$, Fischer's exact test) to context-dependent 5hmC generation, additional factors might exist for TET3 recruitment. In line with this, we identified several other transcriptional regulators interacting with TET3, such as ASXL1, CTCF, MORF4L1, SUZ12, and VAX1. It is tempting to speculate that these transcription regulators could influence TET3 hydroxylase activity in a similar way as REST. For instance, it was shown that the insulator CTCF interacts with TET enzymes

and that depletion of CTCF in mouse ESCs and adipocytes leads to significant loss of 5hmC in transcriptional enhancers (Dubois-Chevalier et al., 2014; Feldmann et al., 2013). However, CTCF seems not to discriminate among the TET isoforms, since it binds to TET1 and TET2 (Dubois-Chevalier et al., 2014) as well as to TET3 (present study).

Taken together, REST and most probably additional DNA-binding transcriptional regulators recruit TET3 to neuronal genes during retinal maturation to oxidize 5mC to 5hmC and activate their transcription. Transcriptional activation goes along with chromatin remodeling and translocation of genomic regions from 5mC-rich heterochromatin to 5hmC-rich euchromatin (Hahn et al., 2013; Mellén et al., 2012). Interestingly, our study reveals that TET3 interacts with several H3K36 methyltransferases. At a functional level TET3 stimulates the activity of NSD3 to generate the activating histone mark H3K36me3. Our data show that the transcriptional repressor REST can facilitate the recruitment of TET3 to 5mC-containing transcriptionally inactive genes (Figure 7D). This repressor-induced and context-specific recruitment is followed by TET3-mediated oxidation of 5mC to 5hmC. Subsequently, H3K36 methyltransferases induce chromatin remodeling to facilitate active transcription (Figure 7D).

Our findings favor a model that involves highly dynamic transcription-factor-dependent DNA hydroxymethylation that supports retinal network maturation by context-specific gene activation. Accordingly, our metabolic feeding experiments showed that the majority of nascent 5hmC marks were generated from pre-existing 5mC, supporting the idea of TET3-induced chromatin remodeling.

5hmC itself is known to recruit specific binding proteins (readers) (Mellén et al., 2012; Spruijt et al., 2013; Takai et al., 2014). We cannot exclude that such 5hmC readers might contribute to the observed TET3-mediated transcriptional activation during retinal maturation. Future study focusing on the identification of developmental-stage-specific 5hmC readers in the retina should help to elucidate their contribution.

In contrast to 5hmC, the levels of the further oxidized methylcytosine derivative 5fC in the retina were very low and decreased during neuronal maturation (see Figure S1C), suggesting that 5fC is a short-lived active DNA demethylation intermediate rather than an independent epigenetic mark in differentiating neuronal cells.

In conclusion, we provide evidence that TET3 and REST play a key role in the activation of neuronal gene transcription. Furthermore, we show that TET3-mediated gene activation involves H3K36 trimethylation-induced chromatin remodeling. This study provides a mechanistic link among REST-directed TET3-mediated 5hmC formation, subsequent chromatin remodeling by NSD3, and gene activation during neuronal maturation.

EXPERIMENTAL PROCEDURES

Animals

All mice used were on the C57BL/6 genetic background. All procedures concerning animals were performed with permission of the local authority (Regierung von Oberbayern). Day of birth was considered as postnatal day 1 (P1).

Lentiviral Constructs and Production

Mouse TET3 was PCR cloned from mouse retinal cDNA using the primers 5'-CTATCTAGAACCGCCATGGACTCAGGGCCAGTGAC-3' (forward) and 5'-TCACCGGTAAGATCCAGCGGCTGTAGGG-3' (reverse), and ligated into a lentiviral vector containing a cytomegalovirus (CMV) promoter (Mistrik et al., 2005), yielding LV-CMV-TET3ΔCXXC-eGFP. LV-CMV-eGFP expressing eGFP only was used for control experiments. Recombinant lentivirus as well as lentiviral particles were generated as described previously (Mistrik et al., 2005).

In Vitro Retinal Explant Cultures

Retinas from wild-type animals were used to generate retinal explants. Animals were killed and the eyes enucleated in DMEM Nutrient Mixture F12-HAM medium (Sigma) containing 10% fetal bovine serum (FBS) and antibiotics. Afterward, the entire eyes were incubated in DMEM-F12 serum-free medium containing 0.12% proteinase K at 37°C for 15 min, to allow preparation of retinal cultures with RPE attached. Subsequently, proteinase K was inactivated using 10% FBS in DMEM Nutrient Mixture F12-HAM medium. The eyes were dissected aseptically in a Petri dish. The anterior segment, lens, vitreous, sclera, and choroids were carefully removed, and the retina was cut perpendicular to its edges, resulting in a cloverleaf-like shape. Subsequently, the retina was transferred to a Millicell culture dish filter insert (Millipore) with the retinal pigment epithelium layer facing the membrane. The insert was put into a six-well culture plate and incubated in DMEM-F12 Nutrient medium at 37°C.

Transduction of Retinal Explant Cultures

Retinal explant cultures from 11-day-old mice were used for transduction experiments using LV-CMV-TET3ΔCXXC-eGFP (neuronal TET3) or LV-CMV-eGFP (control). Twelve retinæ were used for each experiment. The LV particles were added to the scleral and vitreal part of retinal explants directly after preparation of the explant cultures. Every second day, the full volume of DMEM-F12, 1.2 ml per dish, was replaced with fresh medium. Retinal explants were harvested after 12 days in vitro for nuclear protein extraction.

Cell Culture and Transfection

HEK293T cells were grown at 37°C and 10% CO₂ in DMEM supplemented with 10% FBS, 100 units/ml penicillin, and 100 µg/ml streptomycin. The cells were transfected with plasmids expressing mouse REST (Addgene plasmid 21310) and/or mouse TET3 (LV-CMV-TET3ΔCXXC-eGFP, see above) using calcium phosphate. Cells were incubated at 37°C and 10% CO₂ for 48 hr, with an additional medium exchange step after 24 hr from transfection. Upon harvesting, the cells were washed once with PBS after removing the medium, pelleted, and then lysed for subsequent genomic DNA extraction as described below.

For RNAi experiments, REST-specific siRNA duplexes or control siRNA duplexes (OriGene, SR304036) and plasmid-expressing mouse TET3 were transfected using siTRAN (OriGene) in Opti-MEM I reduced serum medium (Invitrogen) according to the manufacturer's instructions. Cells were incubated at 37°C and 10% CO₂ for 48 hr and subsequently harvested for genomic DNA isolation. Knockdown efficiency of gene expression was verified using qRT-PCR.

Tracing Experiments with [¹³C,³D₃]-Labeled L-methionine in Retinal Explants

De novo methylation-tracing experiments with [methyl-¹³C,³D₃]-methionine were performed in p11 retinal explant cultures (as described in the Supplemental Experimental Procedures) that were maintained in methionine-free DMEM containing 10% FBS and supplemented with 2 mM [methyl-¹³C,³D₃]-methionine for 9 days (after 3 days, media were renewed). Four retinal cultures were collected each 3 days after feeding and pooled for subsequent UHPLC-MS/MS quantification.

Data Interpretation of Interactors

From the identified proteins, only the proteins that were enriched in both the forward and the reverse experiments, with an enrichment factor of at least 1.5-fold over the control sample, are claimed as interactors. GO analysis

was performed using Ingenuity pathway analysis (IPA, Ingenuity Systems, <http://www.ingenuity.com/>) or DAVID bioinformatics database (<http://david.abcc.ncifcrf.gov/>).

Acid Extraction of Histones

To examine histone modifications of histone 3 (H3), we followed a protocol from Shechter et al. (2007). In brief, cells were lysed in 1 ml hypotonic buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM KCl, 1.5 mM MgCl₂, and 1 mM DTT. Samples were incubated for 30 min on rotator at 4°C. The intact nuclei were centrifuged at 10,000 × *g* for 10 min at 4°C. Next, the nuclei were re-suspended in 400 μl 0.4 N H₂SO₄ and samples were incubated on rotator for 30 min. The nuclear debris was removed by centrifugation at 16,000 × *g* for 10 min. The supernatant was incubated for 30 min with 132 μl trichloroacetic acid (TCA) to precipitate histones. Histones were pelleted by 16,000 × *g* for 10 min and washed with acetone without disturbing the pellet to remove remaining acid from the solution. The pellets were eluted using ddH₂O.

Quantification of Histone Modifications

The visible protein band of histone H3 was cut out of a Coomassie-stained SDS-gel after destaining. The protein was in-gel alkylated and typically digested (Shevchenko et al., 2006). Trypsin activity was stopped using 1 mM PMSF and the peptides were labeled according to manufacturer's protocol with the TMT 2 (retina) or the TMT 4 (HEK293T cells).

Post-labeled samples of each experiment were pooled together and analyzed via LC-MS/MS. The relative reporter ion intensities of the trimethylated H3K36 peptide were normalized to the relative quantities of the whole protein. Finally, the fold change of the trimethylated peptides was calculated in relation to each control sample.

Bioinformatics

Sequence reads of 5hmC containing DNA fragments were mapped onto the reference mouse genome (NCBI Build UCSC mm9) using the Bowtie (v0.12.7) algorithm. Unique and monoclonal reads were used for further analysis. Refseq genes were downloaded from the UCSC mm9 annotation database (UCSC Genome Browser). GO analyses on differentially hydroxymethylated regions were performed using Ingenuity (IPA, Ingenuity Systems).

ACCESSION NUMBERS

The NCBI GEO accession number for the 5hmC profiling data reported in this paper is GSE65860. The ProteomeXchange accession number for the proteomics data reported in this paper is PXD001857.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and 11 tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.03.020>.

AUTHOR CONTRIBUTIONS

S.M. and T.C. supervised the project. S.M., T.C., and A.P. designed experiments and interpreted data. A.P. conducted the analysis of gene expression, retinal explant culture, and LFQ proteomics. E.S., A.P., and N.M. performed protein pull-down studies. D.E. and S.K.L. performed peptide LC-MS quantification including sample preparation and analysis. A.F.K. conducted LFQ proteomics. M.W., J.S., and A.P. isolated genomic DNA. M.W., A.P., and J.S. performed sample preparation. M.W. and J.S. performed UHPLC-MS/MS quantification and isotope-tracing experiments of DNA bases. A.P., S.K., and S.M. performed and analyzed immunostainings and hMeDIP experiments. S.M. and E.S. generated expression plasmids and lentiviral vectors. A.P. and E.S. performed HEK293 cell culture experiments and acidic histone isolation. V.S. performed PLAs. A.P. and S.M. performed bioinformatics analysis. S.M. wrote the paper with input from other authors. M.B. and M.M. contributed to the experimental design, data interpretation, and manuscript editing.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) cluster of excellence (CIPSM, EXC114). A.F.K. is supported by a fellowship from the Peter und Traudl Engelhorn-Stiftung. We thank Gail Mandel for the gift of anti-REST antibodies; Tim M. Strom and Thomas Meitinger for NGS; Lukas Windhager, Ralf Zimmer, and Michael Bonin for help with bioinformatics analysis; and Fred Koch for help with cloning and qPCR.

Received: November 7, 2014

Revised: January 9, 2015

Accepted: March 7, 2015

Published: April 2, 2015

REFERENCES

- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181.
- Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altschuler, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* 80, 949–957.
- Colquitt, B.M., Allen, W.E., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. USA* 110, 14682–14687.
- Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E., and Hansen, K. (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet.* 8, e1002494.
- Dubois-Chevalier, J., Oger, F., Dehondt, H., Firmin, F.F., Gheeraert, C., Staels, B., Lefebvre, P., and Eeckhoutte, J. (2014). A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res.* 42, 10943–10959.
- Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schübeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* 9, e1003994.
- Gregory-Evans, C.Y., Wallace, V.A., and Gregory-Evans, K. (2013). Gene networks: dissecting pathways in retinal development and disease. *Prog. Retin. Eye Res.* 33, 40–66.
- Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P., and Lu, Q. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep.* 3, 291–300.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Hoon, M., Okawa, H., Della Santina, L., and Wong, R.O. (2014). Functional architecture of the retina: development and disease. *Prog. Retin. Eye Res.* 42, 44–84.
- Huang, Y., Myers, S.J., and Dingleline, R. (1999). Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nat. Neurosci.* 2, 867–872.
- Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Kraner, S.D., Chong, J.A., Tsay, H.J., and Mandel, G. (1992). Silencing the type II sodium channel gene: a model for neural-specific gene regulation. *Neuron* 9, 37–44.

- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., and Gage, F.H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* 116, 779–793.
- Liu, N., Wang, M., Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H., and Spada, F. (2013). Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS ONE* 8, e62755.
- Lu, T., Aron, L., Zullo, J., Pan, Y., Kim, H., Chen, Y., Yang, T.H., Kim, H.M., Drake, D., Liu, X.S., et al. (2014). REST and stress resistance in ageing and Alzheimer's disease. *Nature* 507, 448–454.
- Lunyak, V.V., and Rosenfeld, M.G. (2005). No rest for REST: REST/NRSF regulation of neurogenesis. *Cell* 121, 499–501.
- Mao, C.A., Tsai, W.W., Cho, J.H., Pan, P., Barton, M.C., and Klein, W.H. (2011). Neuronal transcriptional repressor REST suppresses an Atoh7-independent program for initiating retinal ganglion cell development. *Dev. Biol.* 349, 90–99.
- McGann, J.C., Oyer, J.A., Garg, S., Yao, H., Liu, J., Feng, X., Liao, L., Yates, J.R., 3rd, and Mandel, G. (2014). Polycarbonyl- and REST-associated histone deacetylases are independent pathways toward a mature neuronal phenotype. *eLife* 3, e04235.
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430.
- Mistrić, P., Mader, R., Michalakakis, S., Weidinger, M., Pfeifer, A., and Biel, M. (2005). The murine HCN3 gene encodes a hyperpolarization-activated cation channel with slow kinetics and unique response to cyclic nucleotides. *J. Biol. Chem.* 280, 27056–27061.
- Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalakakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* 49, 5375–5377.
- Naruse, Y., Aoki, T., Kojima, T., and Mori, N. (1999). Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proc. Natl. Acad. Sci. USA* 96, 13691–13696.
- Okawa, H., Hoon, M., Yoshimatsu, T., Della Santina, L., and Wong, R.O. (2014). Illuminating the multifaceted roles of neurotransmission in shaping neuronal circuitry. *Neuron* 83, 1303–1318.
- Santiago, M., Antunes, C., Guedes, M., Sousa, N., and Marques, C.J. (2014). TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? *Genomics* 104, 334–340.
- Schiesser, S., Pfaffeneder, T., Sadeghian, K., Hackner, B., Steigenberger, B., Schröder, A.S., Steinbacher, J., Kashiwazaki, G., Höfner, G., Wanner, K.T., et al. (2013). Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* 135, 14593–14599.
- Schoenherr, C.J., and Anderson, D.J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* 267, 1360–1363.
- Shechter, D., Dormann, H.L., Allis, C.D., and Hake, S.B. (2007). Extraction, purification and analysis of histones. *Nat. Protoc.* 2, 1445–1457.
- Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860.
- Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., and Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* 3, 995–1000.
- Solovei, I., Kreysing, M., Lancôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137, 356–368.
- Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29, 68–72.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159.
- Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., Koyama-Nasu, R., Nasu-Nishimura, Y., Katou, Y., Ogawa, H., et al. (2014). 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex. *Cell Rep.* 9, 48–60.
- Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* 15, 69–81.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappaport, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348.
- Xiang, M. (2013). Intrinsic control of mammalian retinogenesis. *Cell. Mol. Life Sci.* 70, 2519–2532.
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42, 451–464.
- Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* 151, 1200–1213.
- Zheng, D., Zhao, K., and Mehler, M.F. (2009). Profiling RE1/REST-mediated histone modifications in the human genome. *Genome Biol.* 10, R9.

3.3 Tet enzymes oxidize T to hmU in mouse embryonic stem cells

Toni Pfaffeneder#, Fabio Spada#, Mirko Wagner#, Caterina Brandmayr, Silvia K. Laube, David Eisen, Matthias Truss, Jessica Steinbacher, Benjamin Hackner, Olga Kotljarova, David Schuermann, Stylianos Michalakis, Olesea Kosmatchev, Stefan Schiesser, Barbara Steigenberger, Nada Raddaoui, Gengo Kashiwazaki, Udo Müller, Cornelia G. Spruijt, Michiel Vermeulen, Heinrich Leonhardt, Primo Schär, Markus Müller, Thomas Carell.
(# geteilte Erstautorenschaft)

Prolog

Die aktive DNA-Demethylierung könnte, wie in Abschnitt 1.6 erläutert, auf vielen möglichen Wegen verlaufen. Ein denkbarer Weg wäre die Desaminierung von 5hmdC zu 5hmdU und anschließende Basenexzisionsreparatur. 5hmdU ist in genomischer DNA aber auch ein oxidativer Schaden von dT der durch reaktive Sauerstoffspezies (ROS) entsteht. In diesem Manuskript wird die Entstehung von 5hmdU in mES-Zellen per quantitativer Massenspektrometrie und Isotopenverfolgung aufgeklärt. Es konnte gezeigt werden, dass der Großteil von 5hmdU in mES-Zellen aus der Tet-vermittelten Oxidation von Thymin stammt.

Autorenbeitrag

Die Isolation der DNA, die massenspektrometrischen Quantifizierung, sowie die Auswertung und Interpretation der Daten führte ich in enger Zusammenarbeit mit *Toni Pfaffeneder und Mirko Wagner* durch. Für eine weitere Auflistung des Beitrags der übrigen Autoren siehe Manuskript.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlages.

Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder^{1,8}, Fabio Spada^{1,8}, Mirko Wagner^{1,8}, Caterina Brandmayr¹, Silvia K Laube¹, David Eisen¹, Matthias Truss², Jessica Steinbacher¹, Benjamin Hackner¹, Olga Kotljarova¹, David Schuermann³, Stylianos Michalakos⁴, Olesea Kosmatchev¹, Stefan Schiesser¹, Barbara Steigenberger¹, Nada Raddaoui¹, Gengo Kashiwazaki¹, Udo Müller⁵, Cornelia G Spruijt⁶, Michiel Vermeulen^{6,7}, Heinrich Leonhardt⁵, Primo Schär³, Markus Müller^{1*} & Thomas Carell^{1*}

Ten eleven translocation (Tet) enzymes oxidize the epigenetically important DNA base 5-methylcytosine (mC) stepwise to 5-hydroxymethylcytosine (hmC), 5-formylcytosine and 5-carboxycytosine. It is currently unknown whether Tet-induced oxidation is limited to cytosine-derived nucleobases or whether other nucleobases are oxidized as well. We synthesized isotopologs of all major oxidized pyrimidine and purine bases and performed quantitative MS to show that Tet-induced oxidation is not limited to mC but that thymine is also a substrate that gives 5-hydroxymethyluracil (hmU) in mouse embryonic stem cells (mESCs). Using MS-based isotope tracing, we show that deamination of hmC does not contribute to the steady-state levels of hmU in mESCs. Protein pull-down experiments in combination with peptide tracing identifies hmU as a base that influences binding of chromatin remodeling proteins and transcription factors, suggesting that hmU has a specific function in stem cells besides triggering DNA repair.

Methylcytosine is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements¹. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation². Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC³, 5-formylcytosine (fC)^{4,5} and 5-carboxycytosine (caC)^{5,6}, whose biological functions are still yet unclear (Fig. 1a)⁷. As fC and caC are both removed by thymine DNA glycosylase (Tdg)^{6,8}, it is currently assumed that they serve as intermediates of an active DNA demethylation process involving base excision repair. In addition to these oxidized C derivatives, cells also contain oxidized T nucleobases such as hmU and fU. These compounds are currently known as oxidative lesions that are thought to form upon the reaction of T with reactive oxygen species (ROS)^{9,10}. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial^{11–14}. Deamination of hmC, situated in a base pair with G (hmC:G), would give rise to hmU:G mismatches, which are known substrates for the DNA glycosylases Tdg, Smug1, Mbd4, Ung2 (ref. 15), Neil1 and Nthl1 (ref. 16). Deamination of hmC:G to hmU:G followed by mismatch repair would therefore establish an alternative pathway to active demethylation (Fig. 1a).

To unravel the origin of oxidized nucleobases, and of hmU in particular, in DNA from mESCs, we performed isotope tracing and quantitative MS studies using the chemically synthesized

isotopologs of mC, hmC, fC, caC, hmU and fU as internal standards (Fig. 1b and Supplementary Results, Supplementary Fig. 1). For the assessment of oxidation products that are formed by the action of ROS, we additionally quantified 8-oxo-G because 8-oxo-G is a well-established ROS reaction product formed from G^{17,18}. We show here that hmU is generated enzymatically from thymidine during stem cell differentiation by the action of the Tet enzymes. A proteomic analysis provides new insight into how genomic hmU can influence the binding of chromatin remodeling proteins and transcription factors.

RESULTS

hmU is present at elevated levels in mESCs

We first created an inventory of the named nucleosides (Supplementary Fig. 1) in mESCs (Fig. 1c) and adult cortex tissue (Fig. 1d). In mESCs, we observed that hmC is, as expected, the most abundant oxidized pyrimidine (10% relative to mC), followed by fC (~1–2% of hmC). The ROS marker 8-oxo-G was detected at similar levels (~45% of fC), showing that nonenzymatic, ROS-induced oxidations of nucleobases are important processes, as expected. We also found relatively high levels of fU (22% of fC). Clearly detectable were also caC and hmU. hmU, which is at the center of this study, was unequivocally detected, as shown by its retention time and its specific fragmentation pattern, which were found to be identical with that of the internal standard [D₂]hmU (Fig. 1e). Notably, both caC and hmU were present in comparable amounts (5% of fC). In adult mouse cortex DNA (Fig. 1d; for other tissues see

¹Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. ²Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. ³Department of Biomedicine, University of Basel, Basel, Switzerland. ⁴Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. ⁵Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. ⁶Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. ⁷Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. ⁸These authors contributed equally to this work.

*e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de

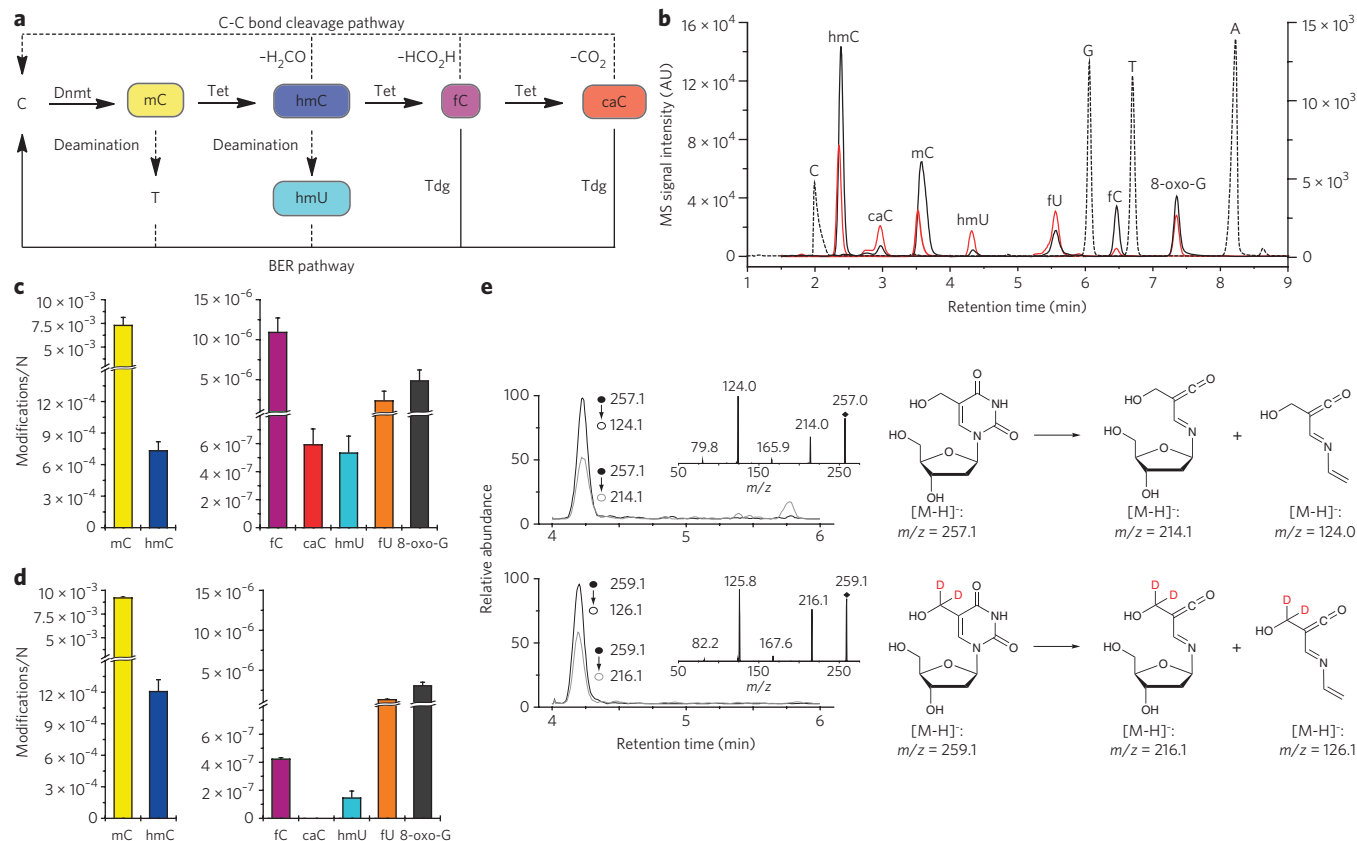


Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV-ESI-MS/MS and levels in mESCs and mouse cortex. (a) Potential active demethylation pathways. (b) Overlaid LC/UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC/UV chromatogram of C, G, T and A is scaled arbitrarily; the overlaid LC/MS/MS chromatograms of hmC, mC, fC and 8-oxo-G are scaled to the left y axis; the LC/MS/MS chromatograms of caC, hmU and fU are scaled to the right y axis. Red chromatograms refer to the corresponding labeled internal standards depicted in **Supplementary Figure 1**. AU, arbitrary units. (c,d) DNA modification levels per nucleoside (N) in mESCs (c; WT01, $n = 7$) and 3-month-old mouse cortex tissue (d; $n = 3$). Depicted are mean values \pm s.d. (e) Representative LC/MS/MS chromatograms for identification and quantification of hmU. Shown are the overlaid chromatograms for the two characteristic fragment ions of hmU (top trace) and the [D₂]hmU internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and [D₂]hmU matching the proposed fragmentation pathway.

Supplementary Fig. 2), the hmC levels are very high (13% relative to mC), as previously reported^{19–21}. In contrast, fC was detected only in very small amounts (0.03% relative to hmC), and caC was not observed at all²². The detected amounts of 8-oxo-G and fU are comparable. Notably, the detected levels of hmU in the cortex were significantly lower compared to that in mESCs (27%; $P = 2.2 \times 10^{-4}$ by Student's t -test). Because the detected levels of 8-oxo-G were similar in the cortex (**Fig. 1d**) and in mESCs (**Fig. 1c**), the elevated hmU levels in mESCs cannot be explained by greater ROS-induced damage in mESCs and must have a different origin.

We next investigated this in more detail and quantified the levels of hmU in three different mESC lines (WT01, J1 and R1) and in a variety of tissues (**Fig. 2a**). Clearly, the hmU levels were higher in mESCs than in somatic tissues by factors of 2–15. The detected hmU levels correspond roughly to 500–1,700 hmU bases per genome in tissue and 2,900–7,800 in mESCs (**Fig. 2a**) depending on the cell type and growth conditions. We cannot explain these differences by elevated oxidative stress levels in mESCs (**Supplementary Fig. 3**). We next analyzed the levels of C, mC and the oxidized pyrimidines hmC, fC, hmU and fU as well as 8-oxo-G in somatic tissues (**Supplementary Fig. 2**) and performed a correlation and cluster analysis to reveal relationships of the modification levels (**Fig. 2b**). The data analysis confirmed that the low levels of hmU in somatic tissues correlate strongly (significant on a level <0.001) with the levels of fU and 8-oxo-G within a separated cluster.

This result showed that in somatic cells, hmU and fU are ROS-induced reaction products. If we assume that in somatic tissues hmU is exclusively formed by ROS, the data showed that in mESCs, 70–80% of the detected hmU is produced by ROS-independent processes (**Supplementary Fig. 3**).

hmU formation by oxidation of T

To analyze the origin of the oxidized nucleobases in mESCs, we performed isotope tracing experiments (**Fig. 3a–c** and **Supplementary Fig. 4**). Substitution of L-methionine with [methyl-¹³CD₃]L-methionine ([¹³CD₃]L-Met) in the growth medium is known to give the labeled S-adenosylmethionine cofactor, which is needed for the conversion of C to mC. Replacement of L-Met by [¹³CD₃]L-Met for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled [¹³CD₃]mC, 88% labeled [hydroxymethyl-¹³CD₂]hmC and 93% labeled [formyl-¹³CD]fC (**Fig. 3b** and **Supplementary Fig. 4**). Within the detection limit (7 and 50 molecules per 10⁸ nucleosides, respectively), no incorporation of the isotopes ¹³C and D into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled [¹³C,¹⁵N₂]thymidine ([¹³C,¹⁵N₂]T) to the growth medium and observed ~76% of label incorporation into T, hmU and fU (**Fig. 3a**, **Supplementary Fig. 4** and **Supplementary Table 1**). The combined data showed that hmU is not generated by deamination of hmC but by oxidation of T (T → hmU; **Fig. 3c**). Consequently, all of the detected hmU

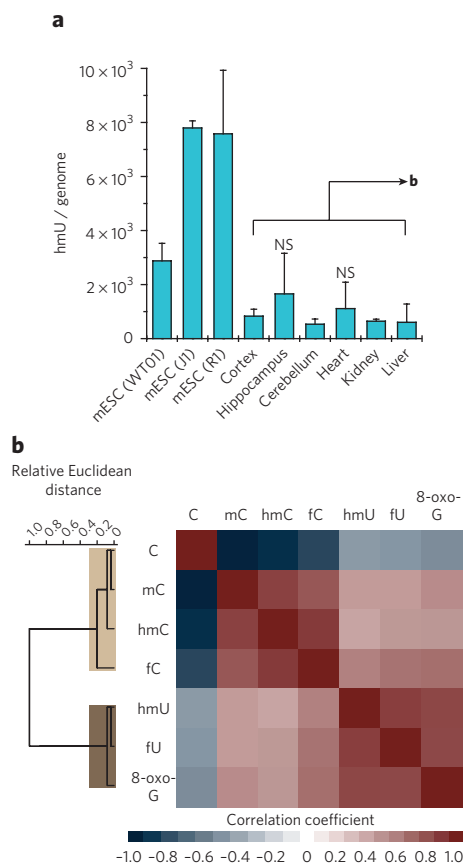


Figure 2 | hmU is present at elevated levels in mESCs compared to tissue.

(a) hmU levels per genome in mESCs ($n_{\text{WT01}} = 7$ replicates, $n_{\text{H}} = 2$, $n_{\text{R1}} = 3$) and mouse tissue (3-month-old individuals, $n = 3$). Levels per genome were obtained considering a mouse genome size of 2.7×10^9 base pairs. Depicted are mean values \pm s.d. The differences between mESCs and mouse tissues are significant ($P = 2.9 \times 10^{-5}$ to 3.6×10^{-2} ; unpaired two-tailed t -test) except for WT01 and hippocampus ($P = 0.292$) or heart ($P = 0.069$). These exceptions are due to higher hmU levels in the hippocampus and heart caused by higher background oxidation (higher 8-oxo-G levels). hmU levels normalized to oxidative background (8-oxo-G) levels are in **Supplementary Figure 3**. NS, not significant. (b) Unsupervised clustering analysis of Pearson correlation coefficients of 24 data sets of selected mouse organs at a defined time point (3-month-old individuals: cortex, hippocampus, cerebellum, heart, liver and kidney). Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001 .

(and also fU) resides in an A base pair context (hmU:A). In this base pair, hmU is repaired by Smug1 but not Tdg¹². This was confirmed by siRNA-mediated knockdown of Smug1 in mESC cells and HEK-293T cells. Indeed, the hmU level increased in these cells (**Supplementary Fig. 5**).

The fact that we were unable to detect hmU derived from hmC deamination, which would be situated in a base pair with G (hmU:G), could also be explained by very fast repair. If repair of the hmU:G base pair is extremely efficient, for example, because deamination and glycosylase-based repair occurs in a tight complex of the involved enzymes, we would be unable to detect this type of hmU because of low steady state levels. Indeed, it was proposed that hmC deamination and hmU excision requires a complex of the cytidine deaminase Aid and the glycosylase Tdg¹². To assess this possibility, we performed isotope tracing experiments using [¹³CD₃]_L-Met in Tdg^{-/-} mESCs stably complemented with either empty vector (control) or a minigene expressing a catalytically incompetent Tdg at near-endogenous levels. These cells are able to form the Aid-Tdg

complex, but the Tdg is inactive, which should give elevated hmU levels if deamination occurs. In both cell lines, we detected high levels of labeled hmC. In the control cells, [hydroxymethyl-¹³CD₂] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [hydroxymethyl-¹³CD₂]hmU was indeed observed, albeit only in small amounts ($\sim 7\%$ of total hmU and $\sim 0.06\%$ of total hmC; **Supplementary Fig. 6**). Thus, the Tdg protein is required for deamination of hmC to hmU. In wild-type (WT) mESCs, the hmU:G mismatches are obviously repaired so quickly that they do not contribute to steady state levels of hmU. All of the detected hmU was derived from T oxidation and resided in hmU:A base pairs.

Tet enzymes form hmU in correlation with mC oxidation

We next investigated whether enzymatic oxidation or ROS-dependent processes are responsible for the observed oxidation of T to hmU. To this end, we studied mESCs with genetic or functional depletions of Tet or DNA methyltransferase (Dnmt) enzymes (**Fig. 4**). Although hmU levels were maintained at normal levels in the severely hypomethylated, Dnmt-depleted cell lines, they were substantially reduced upon knockdown of Tet1 and Tet2. As the levels of fU and 8-oxo-G remained unaffected, the data establish that, though both hmU and fU are generated by T oxidation, their formation must occur by two independent processes in mESCs. The data supported the idea that fU is a ROS-created lesion similar to 8-oxo-G^{9,10,17,18}, whereas most hmU is generated by Tet-induced oxidation of T.

In support of enzymatic T oxidation by Tet proteins, ectopic expression of the catalytic domain of Tet1 (Tet1cd) in HEK-293T cells led to a 65-fold increase for both hmC and hmU levels. This was not observed when a catalytic mutant of Tet1 (Tet1cm) was expressed (**Supplementary Fig. 7a**). To determine whether under these conditions hmU is generated by deamination, as previously suggested^{11,12}, we again replaced natural _L-Met with [¹³CD₃]_L-Met in the medium, but we did not detect incorporation of heavy isotopes into hmU, providing evidence that the elevated hmU levels do not originate from hmC deamination (**Supplementary Table 1**). Finally, we confirmed enzymatic generation of hmU *in vitro* by incubating recombinant Tet1cd with a plasmid that was premethylated by bacterial methyltransferase M.SssI. In addition to oxidation of mC to hmC, fC and caC, we detected hmU at a prominent level of 9% relative to hmC (**Supplementary Fig. 7b**), showing that the catalytic center of the Tet enzymes clearly has the capacity to oxidize T to hmU.

To further confirm that hmU is formed in mESCs in the process of epigenetic reprogramming, we analyzed the dynamic changes of mC and all of the oxidized pyrimidines plus 8-oxo-G during differentiation. It was recently shown that mC and hmC levels sharply increase when mESCs maintained in the naive state are shifted to a primed state in serum-containing medium^{23–25}. To investigate global kinetics of all of the Tet-generated oxidation products under more physiologically relevant priming conditions, we used established protocols based on serum-free N2B27 medium for differentiation of naive mESCs into states resembling that of post-implantation epiblasts^{26,27}. mESCs were first grown for several passages in the presence of MEK and GSK3 inhibitors (dual inhibition or 2i conditions) and LIF to induce a hypomethylated state resembling that of the naive epiblast^{23–25}. The data in **Figure 5a** are averaged from three independent differentiation experiments, each performed with two cell lines in the absence of growth factors. First, we observed that the fU and 8-oxo-G levels stayed constant (**Fig. 5a**), in line with the idea that both are ROS-derived products. In contrast, fC and caC levels peaked at about 8 h. Both hmU and hmC also peak between 8 h and 16 h. Isotope tracing with [¹³CD₃]_L-Met under these conditions provided no evidence for switched-on deamination of hmC to hmU (**Supplementary Table 1**). At these peak levels, we estimated that mESCs contain roughly 110,000 fC bases, 4,400 caC bases and

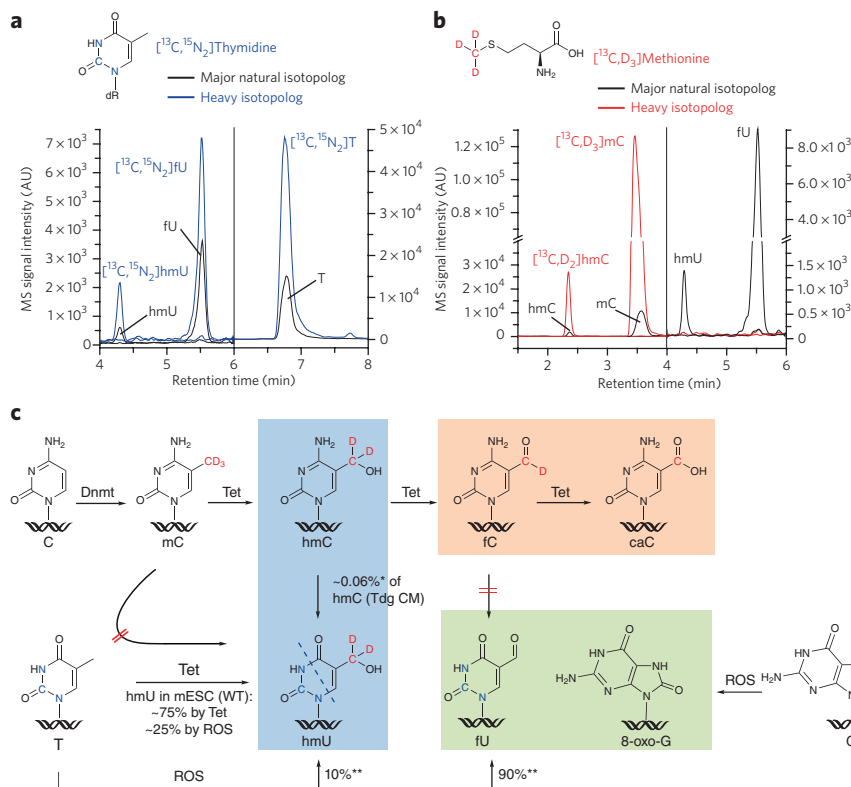


Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination. (a) Overlaid LC/MS/MS chromatograms of heavy ($^{13}\text{C}_5, ^{15}\text{N}_2$)hmU, $^{13}\text{C}_5, ^{15}\text{N}_2$ fU and $^{13}\text{C}_5, ^{15}\text{N}_2$ T; (b) overlaid chromatograms of heavy ($^{13}\text{C}_5, \text{D}_3$)hmC, $^{13}\text{C}_5, \text{D}_3$ fC and $^{13}\text{C}_5, \text{D}_3$ fU; (c) enzymatic and ROS-dependent pathways leading to the formation of hmC, fC, caC, hmU and fU. Single asterisks denote labeled hmU generated by deamination of labeled hmC, which was observed only in Tdg catalytic mutant (CM) cells (Supplementary Fig. 6), representing ~7% of the total hmU content and corresponding to deamination of ~0.06% hmC. Double asterisks denote basal rates of ROS-dependent T oxidation, which were determined in HEK-293T cells, where Tet activity is lowest (3.6 hmC per 10^5 nucleosides; Supplementary Table 2). Here, hmU and fU roughly represent 10% and 90% of T oxidation products relative to the sum of each other.

14,000 hmU bases per genome, showing that, at its peak level, hmU is three times more abundant than caC. The dynamic peaking data allowed us to estimate half-life times for fC (7 h), caC (5 h) and hmU (4 h) during the differentiation process (Supplementary Fig. 8). Using the dynamic quantitative data, we performed a correlation and cluster analysis of the DNA modification levels (Fig. 5b). Three independent clusters were obtained. One cluster involves members of cytosine methylation (and demethylation) dynamics, including C, mC, fC and caC, but, to our surprise, not hmC. Instead, hmC groups with hmU. We therefore concluded that the formation of these modifications is tightly coupled by the action of the Tet enzymes. The ROS-induced lesions fU and 8-oxo-G form the third, well-separated cluster. Notably, hmU does not correlate with the ROS lesions.

Recently, it was shown that exposure of naive mESCs to fibroblast growth factor 2 (FGF-2) and activin A (ActA) for 48 h under similar conditions as those described above gives a homogeneous cell population whose transcriptome closely resembles that of the post-implantation epiblast²⁶. Under these conditions, mC levels increased more rapidly, approaching somatic levels within a time frame closely reflecting that observed during embryonic development (Supplementary Fig. 9a). Levels of hmC rose steadily throughout the 48-h time course. Whereas fC peaked at 36 h, hmU reached its maximum at 24 h (Supplementary Fig. 9b). qPCR data

showed a robust induction of Dnmt3b that is most likely responsible for the increasing mC levels (Supplementary Fig. 9a). A transient peak of Tet1 expression at 24 h went in hand with rising hmU and fC levels, whereas Tet3 expression started rising slowly from the same time point and may drive further accumulation of hmC and fC (Supplementary Fig. 9b). Tdg was progressively upregulated together with the onset of methylation activity, whereas Smug1 showed little, if any, change (Supplementary Fig. 10). These data supported our view that hmU peaking is not caused by downregulation of the repair pathway.

Thus, the kinetic data showed a complex interplay between methylation and two oxidation reactions during differentiation, which depends on the exact conditions. Most important is the fact that hmU showed a time-dependent occurrence similar to the other oxidized bases hmC, fC and caC independently of the conditions investigated, confirming its formation during epigenetic reprogramming.

hmU attracts specific readers

To obtain initial insight into a potential biological function of hmU in comparison to hmC, we screened for specific readers associating with hmU:A as well as with hmC:G and hmU:G using protein pull-down and relative quantification by LC/MS/MS. In previous stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics studies with hmC, fC and caC containing oligonucleotides, we and others observed a high number of specific protein readers, arguing that the new bases influence a variety of different processes^{28,29}. For this hmU study, we further developed this approach for the detection of proteins that directly interact with the modified bases hmU (as well as hmC for comparison) to get a more direct insight into their function (Fig. 6a). For the study, we not

only included into the biotinylated DNA duplexes (24mers) an hmU (hmC) base but also equipped the counter strand with a polyethylene glycol-based linker carrying a reactive *N*-hydroxysuccinimide (NHS) ester moiety and a reductively cleavable disulfide bond in the middle. This reactive linker cannot be inserted into the DNA strand using solid phase synthesis. In addition, the NHS ester does not survive hybridization conditions. We therefore attached the linker as its azide derivative, using Cu(I)-catalyzed click chemistry, to an alkyne-bearing base present in the DNA duplex^{30,31}. These DNA duplexes were subsequently incubated with nuclear extracts from mESCs. The DNA-bound protein complexes were isolated using streptavidin-coated magnetic beads (Fig. 6a)³². DNA duplexes with the canonical base pairs A:T and C:G at the respective positions served as reference strands. The NHS linker has two functions in the experiment: First, it will covalently trap the reader proteins specifically at the ϵ -amino groups of lysines, which allows the identification of transiently binding proteins as well. Second, because the linker is cleavable, it will tag the trapped lysine residues, leaving a defined label on those proteins that bind in close proximity to the modified bases hmC and hmU. This allowed us to distinguish protein readers that bound close to hmC and hmU from proteins that are secondary members of the complexes. The covalently trapped proteins were next tryptically digested and labeled with tandem mass tagging (TMT) isobaric tags to allow protein

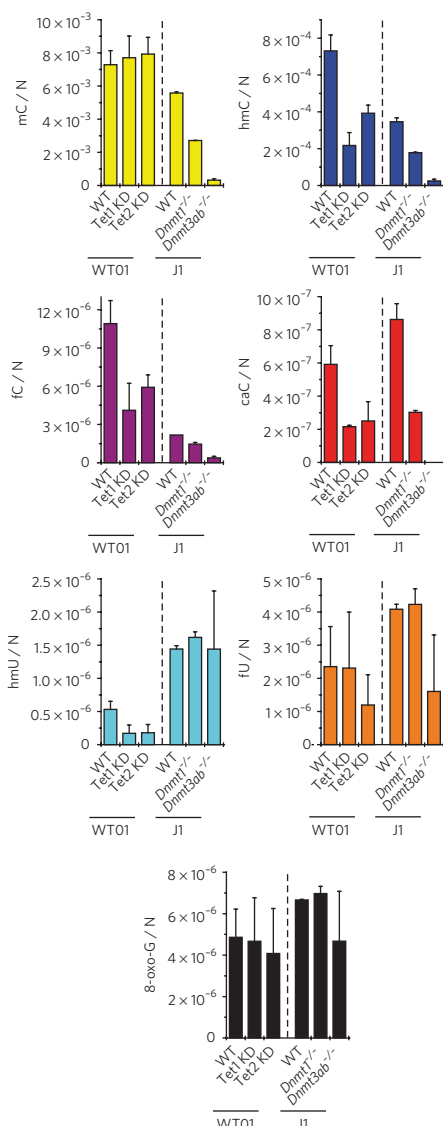


Figure 4 | Tet1 and Tet2 generate hmU in mESCs. Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N) \pm s.d. of biological replicates as follows: wild type (WT01, $n = 7$), Tet1 KD (WT01, $n = 3$) and Tet2 KD (WT01, $n = 3$); WT (J1, $n = 2$), *Dnmt1*^{-/-} (J1, $n = 2$) and *Dnmt3a/b*^{-/-} (J1, $n = 6$).

identification and quantification by MS^{32,33}. Only proteins that were enriched in both the forward and the reverse TMT experiment were considered to be specific hmU (hmC) readers (Fig. 6). All of the readers were subsequently divided into two groups. Proteins that were enriched relative to the control strands are termed specific readers (sRs). Of those, the proteins that were identified with a peptide containing the tag are termed direct-specific readers (dsRs).

The scatter plot (Fig. 6b and Supplementary Fig. 11a) revealed over 250 sRs recognizing the hmU:A base pair (relative to T:A). Of those proteins, 99 were identified with tagged peptides. These proteins were therefore designated dsRs. A gene ontology analysis showed that 64 proteins of the 250 sRs are nucleotide binders, and 25 proteins are involved in chromatin organization (Supplementary Fig. 11b).

Among the dsRs we identified were the regulatory proteins Uhrf2, transcription factor HIVP3 or the poly [ADP-ribose] polymerase 14

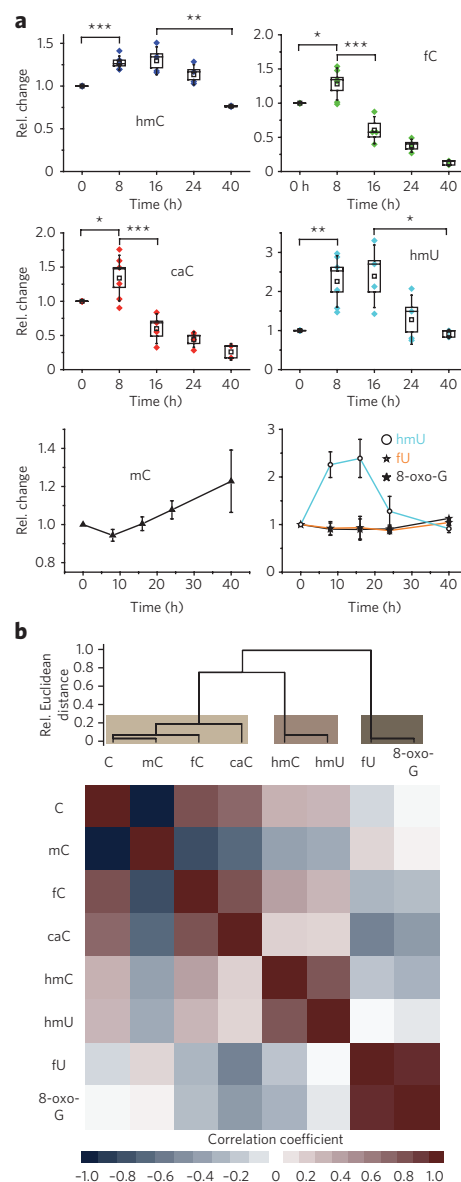


Figure 5 | hmU is produced during mESC differentiation. (a) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using two different mESC lines. In the four upper plots, colored diamonds reflect mean values of technical triplicate measurements, open squares reflect mean values of biological replicates, boxes represent the s.e.m., and whiskers represent the s.d. of the biological replicates; * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.005$ by unpaired two-tailed *t*-test. In the two lower plots, the relative modification changes of mC, hmU, fU and 8-oxo-G are depicted as biological mean values \pm s.e.m. (b) Unsupervised clustering analysis of Pearson correlation coefficients of 22 data sets obtained from the differentiation experiments depicted in a. Discussed correlations are strong to very strong (Pearson coefficient > 0.7) and significant on a level < 0.001 .

(Parp14), which showed that these proteins interacted directly with hmU. We also detected several H2A and H2B histone proteins among the dsRs and the sRs. Notably, in the group of the sRs, we observed also Dnmt3a and Dnmt3b, which are involved in gene expression regulation. Methyl CpG-binding protein 2 (MeCP2) was identified as well, but it showed only low enrichment. We also detected several transcriptional regulators, such as bromodomain PHD finger

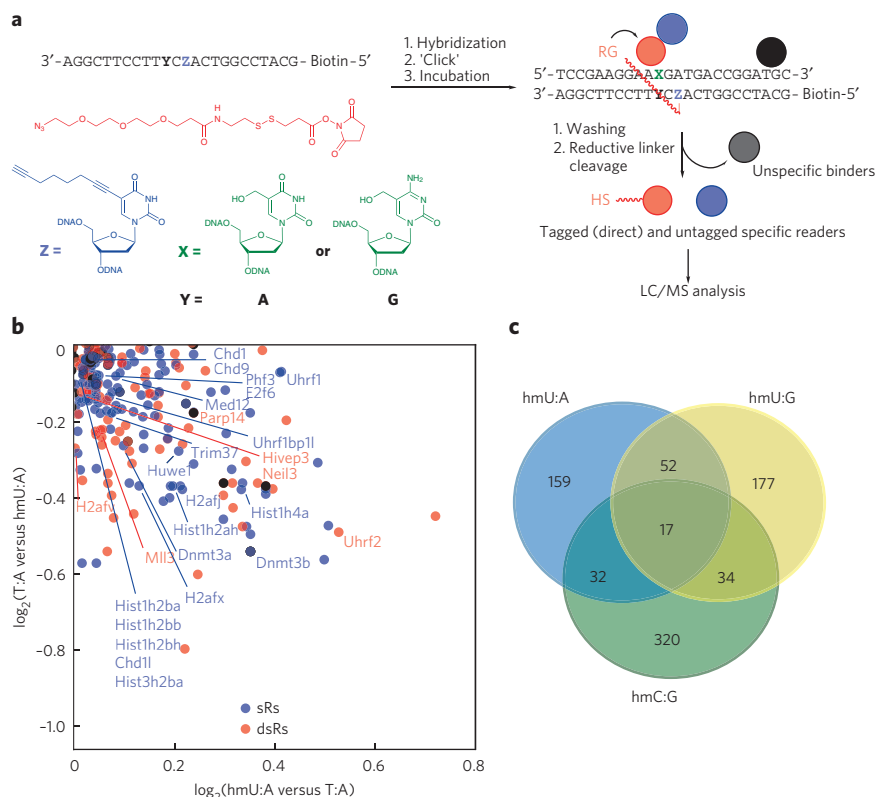


Figure 6 | Identification of hmU:A readers. (a) Workflow of pulldown experiments with hmU:A-containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pulldown are hybridized with DNA oligomers containing hmU (X = hmU and Y = A). After the click reaction with the DNA-protein cross-linker (red), which has an NHS ester as a lysine-specific reactive group (RG), the dsDNA oligomers are incubated with a nuclear extract from mESCs, and specific binders are covalently linked to the DNA. After three washing steps, the linker is reductively cleaved, and the proteins are enzymatically digested and labeled with TMT2plex reagents. The same dsDNA oligomer without hmU modification (X = T and Y = A) was used as a control sample. Specifically enriched proteins are mixed with the control sample before LC/MS analysis. The same workflow was used for experiments with hmU:G or hmC:G and C:G as a control. (b) Magnified view of proteins enriched with hmU:A-containing oligomer. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross-linker and marked in red. Black dots are considered unspecific binders. A full-scale representation is shown in **Supplementary Figure 11**. (c) Venn diagram showing overlap of specific binders for hmU:A, hmU:G and hmC:G.

transcription factor (Bptf), transcription factor E2F6 and mediator of RNA polymerase II transcription subunit 12 (Med12). Moreover, we identified hmU-binding E3 ubiquitin-protein ligases such as Uhrf1, Trim37 and Huwe1 and finally the chromodomain helicase DNA-binding proteins 1 and 9 (Chd1 and Chd9) among the hmU readers, which are chromatin remodeling factors and regulate polymerase I and II transcription. To exclude bias by sequence-specific binding, we repeated the experiment with hmU in a different sequence context. Indeed, 78% of the identified protein readers are also observed in this control experiment, and included among them were all of the proteins discussed above, with the notable exception of Uhrf1, which was not observed in this experiment (**Supplementary Data Set 1**). The data show that hmU:A recruits specific proteins that are involved in chromatin biochemistry in a broad sense.

As a first validation of the MS-identified readers, we focused on Uhrf1 and Uhrf2. Uhrf1 is a known mC and hmC binder³⁴, whereas we recently identified the homolog Uhrf2 as a specific binder of hmC²⁹. The MS data for hmU obtained in this study showed an enrichment of Uhrf1 and Uhrf2. In addition, Uhrf2 was identified as a direct binder. To validate the MS findings, we overexpressed

Uhrf1 or Uhrf2 together with Tet1cd in HEK-293T²⁹ and measured the levels of mC, hmC, fC, caC and also hmU. Indeed, we observed increased levels of hmU, proving a functional relation between hmU levels and Uhrf1 and Uhrf2 (**Supplementary Fig. 12**). As Uhrf2 is known to interact with Dnmt3a and Dnmt3b³⁵ and these two proteins are also identified as sRs of hmU, our initial data provide a consistent picture.

Analysis of the specific hmU readers showed that 49 of the identified proteins were also enriched with an hmC-modified strand. Among these proteins were, for instance, bromodomain PHD finger transcription factor (Bptf), MeCP2 and 19 other nucleotide-binding proteins. The obtained list of specifically binding proteins is in good agreement with our recently published data set²⁹, which further supports the validity of the covalent trapping method. Finally, we compared the results of the hmU:A pulldown with a pulldown of hmU:G, which would be the product of hmC deamination. We found 69 of 280 sRs in both experiments. We identified transcription regulators as sRs as well in the hmU:G pulldown. However, we also found several DNA repair proteins, such as DNA repair protein complementing XP-C cells homolog (XPC), DNA repair protein Rad50 and five helicases, among others (**Supplementary Data Set 1**).

DISCUSSION

This work aimed to analyze the origin of hmU in mESCs. Although hmU as well as fU are both well-characterized oxidation products of thymidine^{9,10}, hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC^{11,12}. We used stable isotopologs of the main oxidation products hmC, fC, caC, hmU, fU and 8-oxo-G in combination with quantitative MS to show that the levels of hmU are strongly elevated in stem cells in comparison to somatic tissue. Notably, we did not observe equally elevated levels of fU and 8-oxo-G, both of which are formed by

the reaction of DNA with ROS. This analysis suggests that hmU in stem cells is not only formed by ROS. The quantitative data allowed us to estimate that in WT stem cells, about 70–80% of the found hmU is not formed by ROS but is derived from a different process. Isotope tracing studies with labeled [¹³CD₃]_L-Met and [¹³C,¹⁵N₂] thymidine showed that the detected hmU originates exclusively from T. Previous studies already proposed a connection between Tet activity and hmU levels, but these studies did not address the origin of hmU, which led to the hypothesis that hmU could form by deamination of hmC^{11,22}. To investigate the involvement of deamination^{11–14}, we studied *Tdg*^{-/-} stem cells reconstituted with a catalytically inactive Tdg. In these cells, the putative deamination–repair complex¹² of Tdg and Aid could form, and we expected to see elevated hmU levels. Indeed, in this experiment, we were able to detect labeled hmU, showing that deamination of hmC does occur, but the levels were low. Most notably, special conditions were required to detect this deamination. In WT cells, all of the detected hmU is clearly derived from T oxidation. As such, the detected hmU is situated in a base pair with A (hmU:A). In conclusion, the majority of hmU in mESCs is produced independently of both ROS and deamination.

Using a combination of functional knockdown in mESCs, ectopic expression of Tet1 catalytic domain in HEK-293T cells and *in vitro* studies with recombinant Tet protein, we showed that the Tet enzymes, known to oxidize mC to hmC, are also responsible for oxidation from T to hmU. A kinetic study in which the temporal changes of all of the oxidized pyrimidines were observed under two different differentiation kinetics showed finally that hmU is formed in processes that also generate hmC and fC. Together with the observed peak levels of hmU of about 14,000 bases per genome, this suggests that hmU could have an epigenetic function similar to hmC. It is conceivable that the Tet enzymes introduce a small amount of hmU during the hmC manufacturing process to trigger faster demethylation by induction of DNA repair. It is known that hmC is not a substrate for repair^{8,12}, whereas hmU is efficiently recognized and repaired³⁶, for example, by the repair glycosylase Smug1 (refs. 37,38). Sporadic introduction of hmU could therefore allow recruitment of repair factors, for example, for long patch repair, as recently suggested³⁹. Alternatively, hmU might influence transcription factor binding, which is known to be a hallmark of epigenetic reprogramming⁴⁰. To study the proteins that closely interact with hmU, a new pulldown experiment was devised in which we inserted a trapping linker into the oligonucleotide. This linker is able to react covalently with proteins that assemble on the DNA duplex close to the hmU base. Because the linker is reductively cleavable, it leaves a tag on the respective lysine residues, which is detected in subsequent MS-based proteomics measurements. Using this technology, we observed that hmU:A recruits transcription factors and proteins that are involved in chromatin biochemistry, including Uhrf1 and Uhrf2. This result is in line with the recent observation that oxidative lesions in DNA can change the binding affinity of the transcription factor CREB when they are present in the respective cognate sequence⁴¹. A further analysis in HEK-293T revealed that the hmU readers Uhrf1 and Uhrf2 are able to modulate the levels of all of the oxidized mC bases and of hmU. It is interesting to note that the observed level increases of hmC were small, whereas they were substantial for hmU and also for the further oxidized bases fC and caC. As Uhrf1 and Uhrf2 are proteins known to be involved in numerous epigenetic processes, our data supported the hypothesis that Tet-induced oxidation of T to hmU may have an epigenetic function. Even in the case that hmU is formed just as a side product of Tet-induced hmC formation, it is now clear that the oxidation chemistry performed by the Tet enzymes has to go in hand with effective DNA repair.

Received 15 December 2013; accepted 17 April 2014;
published online 18 May 2014

METHODS

Methods and any associated references are available in the [online version of the paper](#).

References

- Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
- Franchini, D.M., Schmitz, K.M. & Petersen-Mahrt, S.K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet.* **46**, 419–441 (2012).
- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
- Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
- He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
- Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).
- Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
- Bjelland, S. *et al.* Oxidation of thymine to 5-formyluracil in DNA: mechanisms of formation, structural implications, and base excision by human cell free extracts. *Biochemistry* **34**, 14758–14764 (1995).
- Mouret, J.F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
- Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
- Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
- Rangam, G., Schmitz, K.M., Cobb, A.J. & Petersen-Mahrt, S.K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS ONE* **7**, e43279 (2012).
- Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
- Zhang, Q.M. *et al.* DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair (Amst.)* **4**, 71–79 (2005).
- Burrows, C.J. Surviving an oxygen atmosphere: DNA damage and repair. *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 147–156 (2009).
- Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat. Protoc.* **3**, 1287–1298 (2008).
- Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
- Liu, S. *et al.* Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res.* **41**, 6421–6429 (2013).
- Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
- Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
- Leitch, H.G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
- Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).
- Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
- Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
- Burley, G.A. *et al.* Directed DNA metallization. *J. Am. Chem. Soc.* **128**, 1398–1399 (2006).
- Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
- Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003); erratum **75**, 4942 (2003); erratum **78**, 4235 (2006).
- Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
- Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
- Pichler, G. *et al.* Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* **112**, 2585–2593 (2011).
- Lewis, H.L., Muhleman, D.R. & Ward, J.F. Serologic assay of DNA base damage. I. 5-Hydroxymethyldeoxyuridine, a radiation product of thymidine. *Radiat. Res.* **75**, 305–316 (1978).
- Boorstein, R.J. *et al.* Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).

38. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H.E. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst.)* **6**, 505–516 (2007).
39. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**, 39 (2013).
40. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
41. Moore, S.P.G., Toomire, K.J. & Strauss, P.R. DNA modifications repaired by base excision repair are epigenetic. *DNA Repair (Amst.)* **12**, 1152–1158 (2013).

Acknowledgments

We thank the Excellence Cluster Center for Integrated Protein Science Munich (CiPS^M) and the collaborative research centers SFB749, SFB646 and SFB1032 as well as German Research Foundation (DFG) grant CA275/8-4, the Volkswagen foundation, NGFNplus (01GS0870) and the Netherlands Organization for Scientific Research (NWO-VIDI) for financial support. T.P. and S.S. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for a predoctoral fellowship. G.K. thanks the Japan Society for the Promotion of Science (JSPS) for a postdoctoral fellowship for research abroad. We thank M. Moser (Max Planck Institute for Biochemistry) for providing R1- and C57Bl6/129-derived mESCs, G. Höfner and K.T. Wanner for their initial help with MS as well as M. Wirsing, L. Belzner and P. Laube for providing bioinformatic tools for data processing.

Author contributions

T.P. synthesized MS standards, performed the sample preparation and ultra high-performance LC/MS/MS method development, did LC/MS analysis, interpreted data and performed statistical analysis. F.S. and N.R. performed mESC differentiation and isotope tracing experiments. M.W. and C.B. performed the HEK-293T experiments, did LC/MS analysis and interpreted data. C.B. performed qPCR and analyzed the data. S.K.L. and D.E. performed the protein pulldown studies and interpreted data. M.T. performed mESC knockdown and knockout experiments. J.S. and O. Kosmatchev did sample preparation and LC/MS analysis. B.H., S.S. and J.S. prepared MS standards. O. Kotljarova performed *in vitro* assays. B.S. synthesized oligonucleotides for protein capture. G.K. synthesized tandem mass tags, and S.M. provided mouse tissue samples. U.M. and H.L. constructed Tet expression plasmids. C.G.S. and M.V. performed Uhrf1/2 overexpression in HEK-293T cells. P.S. and D.S. provided plasmids and cell lines. M.M. and T.C. conceived and supervised the project, interpreted data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to M.M. and T.C.

ONLINE METHODS

General materials and methods. Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR or Acros Organics and used without further purification. Acetonitrile of LC/MS grade was purchased from Carl Roth GmbH & Co., KG. Formic acid, p.a. for MS, was purchased from Fluka, and water was purified with a Milli-Q Plus system from Merck Millipore.

The MS standards 5-methyl-2'-deoxycytidine (mC), 5-trideuteromethyl-2'-deoxycytidine ($[D_3]$ mC), 5-hydroxymethyl-2'-deoxycytidine (hmC), 5-dideuterohydroxymethyl-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -cytidine ($[^{15}N_2,D_2]$ hmC), 5-formyl-2'-deoxycytidine (fC), 5-formyl-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -cytidine ($[^{15}N_2]$ fC), 5-carboxy-2'-deoxycytidine (caC), 5-carboxy-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -cytidine ($[^{15}N_2]$ caC), 5-hydroxymethyl-2'-deoxyuridine (hmU), 5-(dideuterohydroxymethyl)-2'-deoxyuridine ($[D_2]$ hmU), 5-formyl-2'-deoxyuridine (fU) and 5-formyl-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -uridine ($[^{15}N_2]$ fU) were synthesized according to earlier published work^{19,21,42}. All of the synthesized compounds were characterized and purity confirmed by 1H -NMR, ^{13}C -NMR and ESI-MS, and some were additionally validated by ^{15}N -NMR. 8-hydroxy-2'-deoxy- $(^{15}N_5)$ -guanosine ($[^{15}N_5]$ 8-oxo-G) (99 atom% ^{15}N) was purchased from Cambridge Isotope Laboratories; 8-hydroxy-2'-deoxyguanosine (8-oxo-G) was from BIOLOG; 2'-deoxyguanosine (G) and 2'-deoxycytidine (C) were from ChemGenes. (methyl- $^{13}C,D_3$)-L-Met (99 atom% D and ^{13}C) was purchased from Sigma-Aldrich, and 2'-deoxy- $(C^2-^{13}C, N^1,N^3-^{15}N_2)$ -thymidine (99 atom% ^{13}C and ^{15}N) from Hartmann Analytic. Aqueous stock solutions of these compounds were stored at $-20^\circ C$ and warmed up to RT before usage.

Oligonucleotide synthesis. Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions (DMT off) and acetyl-protected dC. Phosphoramidites (including the 5-hydroxymethyl-dU-CE phosphoramidite and the 5'-biotin phosphoramidite) and polystyrene carriers were obtained from Glen Research. 5-octadynyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by 1H -NMR, ^{13}C -NMR, ^{31}P -NMR and ESI-MS⁴³. The crude oligonucleotide was cleaved from the resin and deprotected in 30% (v/v) ammonium hydroxide and 40% (v/v) methylamine (1:1) at $65^\circ C$ for 10 min and purified by preparative and analytical HPLC (Waters Breeze and Alliance, respectively). Separation was performed by applying a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 ml/min) from Macherey-Nagel with a gradient of buffer A (0.1 M NH_4Et_3OAc in water) and buffer B (0.1 M NH_4Et_3OAc in 80% MeCN). DNA-containing fractions were characterized by MALDI-TOF (Bruker Autoflex II) and analytical HPLC, combined, and desalted by C18-Sep-Pak cartridges (Waters). For analytical HPLC, separation was performed by applying a CC 250/4 Nucleosil 120-3 C18 column from Macherey-Nagel (flow: 0.5 ml/min) with the aforementioned buffer system.

For the protein pulldown studies, an hmU-containing oligonucleotide was hybridized with a DNA strand modified with a 5'-biotin and a 5-octadynyl-dU. The sequences of the DNA strands are summarized in **Supplementary Table 3**. To 10 nmol of the dsDNA, 0.5 μ l of a 200 mM solution of the cross-linking azide-PEG₃-S-S-NHS ester ($C_{18}H_{29}N_5O_8S_2$, Jena Bioscience, Jena, Germany) in DMSO was added. In a separate tube, CuBr was dissolved in a TBTA solution (DMSO/*t*BuOH 3:1, 100 mM) resulting in a 1:1 Cu(I):TBTA ratio. This solution was immediately added to the DNA/azide mixture. Furthermore, 45 μ l of DMSO/*t*BuOH (3:1) were added, the mixture was shaken at $37^\circ C$ for 3 h, and the resulting product was purified by ethanol precipitation.

LC/MS analysis of DNA samples. Quantitative LC/UV-ESI-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. An improved method, based on earlier published work^{29,42,44,45}, was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature $50^\circ C$, gas flow 15 l/min (N_2), nebulizer 30 psi, sheath gas heater $275^\circ C$, sheath gas flow 11 l/min (N_2), capillary voltage 2,500 V in the positive ion mode, capillary voltage $-2,250$ V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 4** and **5**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μ m, 2.1 mm \times 150 mm) at $30^\circ C$ using a gradient of water

and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min: 0 \rightarrow 5 min; 0 \rightarrow 3.5% (v/v) MeCN; 5 \rightarrow 6.9 min; 3.5 \rightarrow 5% MeCN; 6.9 \rightarrow 7.2 min; 5 \rightarrow 80% MeCN; 7.2 \rightarrow 10.5 min; 80% MeCN; 10.5 \rightarrow 11.3 min; 80 \rightarrow 0% MeCN; 11.3 \rightarrow 13 min; 0% MeCN. The effluent up to 1.5 min and after 9 min was diverted to waste by a Valco valve. The autosampler was cooled to $10^\circ C$. The injection volume was amounted to 29 μ l. Calibration curves, method validation and data processing are in **Supplementary Note 2**. A complete compilation of LC/MS quantifications results see **Supplementary Note 3**.

DNA digestion. 5–25 μ g of genomic DNA in 25 μ l H_2O were digested as follows: An aqueous solution (7.5 μ l) of 480 μ M $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards (**Supplementary Note 2**) were added, and the mixture was incubated at $37^\circ C$ for 3 h. After addition of 7.5 μ l of a 520 μ M $[Na]_2$ -EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h at $37^\circ C$ and then stored at $-20^\circ C$. Prior to LC/MS/MS analysis, samples with up to 15 μ g DNA, for which the quantification of low amounts of caC was aspired, were filtered by using an AcroPrep Advance 96 filter plate 0.2 μ m Supor (Pall Life Sciences). In contrast, samples with 15–25 μ g DNA (isotope-tracing experiments) were filtered by using an AcroPrep Advance 96 filter plate 10K Omega (Pall Life Sciences).

Genomic DNA isolation. Tissues of female WT mice (C57-BL/6N) were dissected at postnatal day 90 and prepared as earlier described^{19,21}. Genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit except for mESC samples differentiated in the presence of growth factors (see below). Extraction was performed following the manufacturer's instructions for genomic DNA isolation from cell culture samples or tissue samples, respectively. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 μ M) and deferoxamine mesylate salt (desferal, 200 μ M) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 μ M), according to published methods, to reduce background oxidation or deamination¹⁸. Elution buffer QF was supplemented with 200 μ M BHT. Following elution, all steps were performed on ice. DNA was then precipitated with NaOAc (0.3 M final) and 0.7 volumes *i*PrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H_2O containing 20 μ M BHT using a Qiagen TissueLyser (30 Hz, 2 min). DNA pellets from mouse tissues were resuspended in PBS buffer and additionally extracted with phenol/ $CHCl_3$, precipitated, washed and resuspended as described above.

R1 mESC samples differentiated in the presence of growth factors or transfected with Smug1 esiRNAs were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was repeated four times with 100 μ l of elution buffer supplemented with BHT (200 μ M). Eluted DNA was precipitated with 2 M ammonium acetate and two volumes of absolute ethanol and finally resuspended in H_2O containing 20 μ M BHT. The flow-through from the spin columns was used to isolate RNA (see real-time PCR analysis).

mESC cell culture. Feeder independent WT01 mESCs (C57BL/6 strain)⁴⁶ were cultured in the presence of serum and LIF as previously described⁴. *Tdg*^{+/−} and *Tdg*^{−/−} mESCs were described previously⁴⁷. *Tdg*^{−/−} mESCs were complemented by random integration of either empty vector (hereafter referred to as *Tdg*^{−/−} mESCs) or a minigene expressing catalytically inactive Tdg (N151A)⁴⁷. Clonal mESC lines with targeted Tdg alleles, R1 cells (strain 129/Sv)⁴⁸, J1 cell lines (strain 129S4/SvJae)⁴⁹ and a mESC line derived from C57BL/6/129 mixed background⁵⁰ were routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1 \times MEM-nonessential amino acids (NEAA), 0.2 mM L-alanyl-L-glutamine, 100 U/ml penicillin, 100 μ g/ml streptomycin (all from PAA), 0.1 mM β -mercaptoethanol, 20 ng/ml ($\geq 1 \times 10^3$ U/ml) mouse recombinant LIF (ORF Genetics), 1 μ M PD 0325901 and 3 μ M CHIR 99021 (2i; both from Axon Medchem). In these conditions, the global levels of genomic mC were very low (and, as a consequence, the levels of its oxidized derivatives were even lower; data not shown). Before DNA isolation, 2i cultures were passaged twice (over 5 d) in DMEM supplemented with FBS and LIF as above but lacking 2i. With this strategy, primed mESC cultures were obtained with no sign of overt differentiation and modified genomic cytosines reached reproducibly higher and stable levels. For isotope tracing with heavy

thymidine in serum-primed mESCs 2i cultures of R1, cells were passaged twice (5 d) in the same serum-containing medium lacking 2i and simultaneously supplemented with 100 μ M [^{13}C , $^{15}\text{N}_2$]T. For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of R1 cells and mESC lines with targeted Tdg alleles were passaged twice (over 5 d) without 2i in L-Met-free DMEM (Life Technologies) supplemented as above and with 0.2 mM of either [methyl- ^{13}C , $^{15}\text{N}_2$]L-Met or natural L-Met. For mESC differentiation without growth factors, R1 cells and the C57Bl/6/129 mixed background cell line were first plated at 1×10^5 cells/cm 2 on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival⁵¹. After 12 h, the medium was replaced without addition of LIF (defined as time point 0 h). The medium was replaced once more at 24 h. For isotope tracing with [methyl- ^{13}C , $^{15}\text{N}_2$]L-Met during mESC differentiation, R1 cells were cultured for two passages in L-Met-free N2B27 medium supplemented with LIF, 2i and 0.2 mM of either [methyl- ^{13}C , $^{15}\text{N}_2$]L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described²⁶, with minor modifications. Briefly, mESCs were cultured in N2B27 medium containing 2i and 1,000 U/ml LIF for several passages and then seeded at 2.2×10^5 cells/cm 2 in N2B27 medium containing 1% KnockOut Serum Replacement (Life Technologies), 12 ng/ml FGF-2 (PeproTech) and 20 ng/ml ActA (ORF Genetics) on plates coated with a thin layer of Geltrex extracellular matrix preparation (Life Technologies). The medium was exchanged after 24 h.

Knockdown experiments in mESCs. shRNA expression vectors targeting Tet1 and Tet2 were generated by cloning synthetic oligonucleotides in pLKO.1 (ref. 52). Recombinant lentiviruses were produced by cotransfecting pLKO.1 shRNA expression vectors and packaging plasmids in HEK-293 cells. 48 h after transduction in the presence of 8 μ g/ml polybrene, shRNA-expressing mESCs were selected with 4 μ g/ml puromycin. Cell pools were continuously cultured in the presence of puromycin. shRNA target sequences were as follows: SCR (control), 5'-CCT AAG GTT AAG TCG CCC TCG-3' (ref. 52); Tet1, 5'-TGT AGA CCA TCA CTG TTC GAC-3' (see ref. 52); Tet2, 5'-TTC GGA GGA GAA GGG TCA TAA-3'. esiRNAs for Smug1 knockdown were generated as described⁵³. The cDNA template for *in vitro* transcription was generated by PCR using following primers: forward, 5'-CGT AAT ACG ACT CAC TAT AGG GAG CCC GTG GGT G-3', and reverse, 5'-CGT AAT ACG ACT CAC TAT AGG GGT TTC GTC CAC TGG G-3'. R1 mESCs were weaned from 2i for two passages in FBS- and LIF-containing medium as described above. Upon plating the second passage, the cells were transfected in a p60 plate with 6 μ g of Smug1 esiRNAs (34.5 nM) and 20 μ l of Lipofectamine RNAi MAX (Life Technologies) according to the manufacturer's instructions and were lysed 72 h after transfection.

Culture and transfection procedures for HEK-293T cells. All transfections were performed using jetPRIME transfection reagent (PEQLAB Biotechnologie GmbH) according to the manufacturer's instructions. HEK-293T cells were seeded 24 h before transfection at a density of 2.5×10^6 cells per 75 cm 2 flask and incubated in 10 ml of medium. The transfection solution (500 μ l of jetPRIME buffer, a specific amount of plasmid DNA (Supplementary Note 4) and 20 μ l of jetPRIME reagent) was added to the medium, and the cells were incubated for 48 h, with an additional medium exchange 24 h after transfection. When cotransfection of esiRNA was performed, a second transfection step (500 μ l of jetPRIME buffer, 5 μ g of esiRNA and 20 μ l of jetPRIME reagent) was carried out 4 h after transfection of plasmid DNA. esiRNAs were purchased from Sigma (human TDG esiRNA EHU038971; human SMUG1 esiRNA EHU098861; human CDK5RAP1 esiRNA EHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

Isotope tracing with [^{13}C , $^{15}\text{N}_2$]thymidine or [^{13}C , $^{15}\text{N}_2$]L-Met in HEK-293T cells transfected with Tet1cd. 24 h before transfection, 2.5×10^6 cells were seeded in a 75-cm 2 flask containing 10 ml either of (for [^{13}C , $^{15}\text{N}_2$]thymidine) DMEM medium supplemented with 50 μ M [^{13}C , $^{15}\text{N}_2$]T or (for [^{13}C , $^{15}\text{N}_2$]L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM [^{13}C , $^{15}\text{N}_2$]L-Met and 0.2 mM L-cystine. Transfection was performed as described above using labeled medium.

Real-time PCR analysis of mRNA expression. For analysis of Tet1 and Tet2 knockdown in mESC total RNA was prepared with Trizol (Invitrogen), cDNA synthesis was performed with Quantitect reverse transcription kit from Qiagen, and real-time PCR was performed with the Power Sybr Green PCR master mix from Applied Biosystems on an Applied Biosystems 7500 Fast

system. Knockdown efficiencies relative to control samples transfected with SCR esiRNAs were 79% and 70% for Tet1 and Tet2, respectively. The primers used to estimate them are listed in **Supplementary Note 4**. For analysis of Smug1 knockdown in mESCs and EpiLC differentiation samples, total RNA was prepared with RNeasy spin columns (Qiagen), followed by DNase treatment using TURBO DNA-free (Ambion, Life Technologies); cDNA synthesis was carried out using iScript cDNA Synthesis kit (Bio-Rad); real-time PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad). Smug1 knockdown efficiency relative to control samples transfected with esiRNAs targeting GFP was estimated to be 60%. Quantification of Tet, Dnmt, Tdg and Smug1 transcripts during EpiLC differentiation and Smug1 knockdown samples was performed using the primers listed in **Supplementary Note 4**. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h.

Tet *in vitro* assay. A plasmid was prepared from *dam*⁻/*dcm*⁻ competent *E. coli* strain (New England BioLabs) and methylated with M.SssI (New England BioLabs). 1.5 μ g of plasmid DNA were then treated with recombinant Tet1 from the 5hmC TAB-Seq Kit (Wisegene) corresponding to ref. 54. After 3 h incubation at 37 °C and proteinase K treatment, the oxidized plasmids were purified with GeneJET PCR Purification Kit from Thermo Scientific and eluted in 25 μ l water. Samples were then subjected to LC/MS/MS analysis as described⁴². The results are compiled in **Supplementary Note 3**.

Correlation and cluster analyses. Statistical data analysis was performed using IBM SPSS Statistics 19. Results of bivariate correlation analyses are summarized in **Supplementary Note 5**. Unsupervised clustering of species with respect to its correlation coefficients was applied by average linkage hierarchical clustering using a squared Euclidean distance measure.

Pulldown assay. For the pulldown assay, 250 μ g (50 μ l) of the crude nuclear protein extracts were filled up to 500 μ l with 50 mM TEAB and 1 mM MgCl $_2$. The binding conditions were 45 mM TEAB, 1.1 mM MgCl $_2$, 2 mM HEPES, 42 mM NaCl and 20 μ M EDTA. Complete Protease Inhibitor Cocktail Tablets were used from Roche Diagnostics (Indianapolis, IN, USA). The DNA oligomers with DNA-protein cross-linker were dissolved in neat DMSO, and 1 nmol was added to the protein lysate and incubated for 20 min at room temperature. Streptavidin-coated magnetic particles (Roche Diagnostics, Indianapolis, IN, USA) were washed three times with binding buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 7.4) before 200 μ l of the bead slurry (equal to 2 mg beads) were added to the sample. Following 2 h incubation at room temperature under constant rotating, the beads were washed three times with 50 mM TEAB and 1 mM MgCl $_2$. The beads were reconstituted in 50 mM TEAB and 1 mM MgCl $_2$. Disulfide bonds of the cross linker were cleaved and alkylated in the process of enzymatic digestion, and the magnetic particles were removed before adding trypsin (described below).

Protein sample preparation. Cell lysis of mouse embryonic stem cells was performed as described in ref. 29. For each lysis, approximately 7.5×10^7 cells were used. Protein concentration was determined by Bradford assay. For each experiment 250 μ g (50 μ l) of the crude nuclear protein extract were used. Protein samples for MS analysis were reduced by adding 100 mM TCEP and by incubating on a shaker at 650 r.p.m. for 1 h at 60 °C and subsequently alkylated with 200 mM iodoacetamide in the dark for 30 min at 25 °C. Protein samples were digested with 0.5 μ g trypsin (Promega, Madison, MA, USA) for 16 h at 37 °C. The reaction was stopped using 1 mM phenylmethylsulfonylfluoride. After tryptic digestion, peptide labeling with the TMT2plex reagents (Thermo Fisher Scientific, Waltham, MA, USA) was performed according to the manufacturer's instructions. TMT2plex reagents 126 and 127 were used to label the samples. When the sample (proteins enriched with hmU- or hmC-containing DNA strands) was labeled with TMT126, the control sample (proteins enriched with no modified DNA-bases) was labeled with the TMT127 reagent and vice versa. Subsequent to the labeling, both sample and control, were combined. This way, each experiment was performed twice as a so-called label swap experiment. Organic solvent was removed by vacuum centrifugation, and the sample was finally reconstituted in 1% (v/v) formic acid for MS analysis.

LC/MS analysis of protein samples. The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific, Waltham, MA, USA) coupled to an LTQ-Orbitrap XL (Thermo Fisher

Scientific, Waltham, MA, USA). Of each eluate, 15 μ l were injected for the analysis. The samples were desalted and concentrated on a μ -precolumn cartridge (PepMap100, C18, 5 μ M, 100 \AA , size 300 μ m i.d. x 5 mm) and further processed on a custom-made analytical column (ReproSil-Pur, C18, 3 μ M, 120 \AA , packed into a 75 μ m i.d. x 150 mm and 8 μ m picotip emitter). A 57-min multistep analytical separation was performed at a flow rate of 300 nl/min. In the first 50 min, a linear gradient was ramped up from 5% (v/v) solvent B (acetonitrile containing 0.1% formic acid and 5% DMSO) and 95% solvent A (water containing 0.1% formic acid and 5% DMSO) to 95% solvent B. This level was held for 5 min and then ramped down again to 5% solvent B within 2 min. Mass spectrometric analyses were performed starting with a full mass scan in the mass range between m/z 300 and m/z 1,650. This survey scan was followed by three MS/MS scans using the FTMS mass analyzer and high normalized collision energy of 70 in the HCD cell and three additional scans using the ion trap mass analyzer and a normalized collision energy of 35.

Protein identification and relative quantification method. The Thermo Proteome Discoverer 1.1 software (Thermo Fisher Scientific, Waltham, MA, USA) was used for protein identification and for relative quantification. The Sequest (Thermo Fisher Scientific, Waltham, MA, USA) search engine was used in combination with a Uniprot database (*Mus musculus*; date of download, 04/2013). As a limit of detection, a ratio of threefold signal over the noise filter was applied. A maximum of two missed cleavage sites was allowed. The mass tolerances were 10 p.p.m. for the precursor mass and 0.5 Da for the fragment ion mass. Carbamidocysteine was set as static modification. Dynamic modifications were: cation, Na (D, E); the residue of the DNA-protein crosslinker (+146.028 Da; K, Y); Oxidation (M) as well as TMT2plex (N-term. and K). Identified, nonredundant peptides, which were labeled with the TMT2 reagent, were used for relative quantification. The integration window tolerance was 20 p.p.m., and the integration method was set to 'most confident centroid'. The signals of the TMT2 reporter ions 126 and 127 were used to calculate ratios and monitor either preferred or nonpreferred binding of the identified proteins to the modified DNA bases in comparison to the control strand. From the identified

proteins, the only proteins considered as 'specific readers' were enriched in both the forward and the reverse experiment.

42. Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
43. Gierlich, J. *et al.* Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
44. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
45. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
46. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
47. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
48. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
49. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
50. Montanez, E. *et al.* Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev.* **22**, 1325–1330 (2008).
51. Ying, Q.L. & Smith, A.G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
52. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
53. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
54. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

3.4 Synthese eines DNA-Promotorsegments mit allen vier epigenetischen Nukleosiden: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl- und 5-Carboxy-2'-Desoxycytidin

Arne S. Schröder, Jessica Steinbacher, Barbara Steigenberger, Felix A. Gnerlich, Stefan Schiesser, Toni Pfaffeneder, Thomas Carell*

Prolog

In diesem Manuskript wurde ein 5-Formyl-2'-desoxycytidin(5fdC)-Phosphoramidit-Baustein entwickelt, der die Synthese fdC-haltiger DNA in ausgezeichneter Reinheit und Ausbeute ermöglicht. In Kombination mit Phosphoramiditen für 5-Methyl-dC, 5-Hydroxymethyl-dC und 5-Carboxy-dC gelang die Herstellung eines Segments des OCT4-Promotors, das alle vier epigenetischen Basen enthält. Aufgrund des großen Interesses an diesen neuen epigenetischen Basen sollte die Fähigkeit, alle vier von ihnen in DNA einzubringen, von besonderer Bedeutung sein.

Autorenbeitrag

Die Synthese der Phosphoramidit Bausteine sowie die Optimierung und Durchführung der DNA-Synthese wurde von Arne Schröder durchgeführt.

Der Verdau der synthetischen DNA, die Entwicklung einer geeigneten Messmethode sowie die massenspektrometrische Quantifizierung und die Auswertung der Daten wurde von mir vorgenommen. Für eine weitere Auflistung des Beitrags der übrigen Autoren siehe Manuskript.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlages.



Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-Deoxycytidine**

Arne S. Schröder, Jessica Steinbacher, Barbara Steigenberger, Felix A. Gnerlich, Stefan Schiesser, Toni Pfaffeneder, and Thomas Carell*

Abstract: A 5-formyl-2'-deoxycytidine (fdC) phosphoramidite building block that enables the synthesis of fdC-containing DNA with excellent purity and yield has been developed. In combination with phosphoramidites for 5-methyl-dC, 5-hydroxymethyl-dC, and carboxy-dC, it was possible to prepare a segment of the OCT-4 promoter that contains all four epigenetic bases. Because of the enormous interest in these new epigenetic bases, the ability to insert all four of them into DNA should be of great value for the scientific community.

Aside from the four canonical nucleosides, the genome contains several DNA modifications that are of epigenetic importance. The nucleoside 2'-deoxycytidine is often methylated at the C5 position to give 5-methyl-2'-deoxycytidine (**1**, mdC), which is involved in epigenetic gene silencing.^[1] In recent years, the oxidation products of mdC were discovered.^[2–5] Ten-eleven-translocation proteins (TET1–3) are now known to oxidize mdC to 5-hydroxymethyl- (**2**, hmdC), 5-formyl- (**3**, fdC), and 5-carboxy-2'-deoxycytidine (**4**, cadC).^[6,7] These “new” modifications (Figure 1 A) are primarily found in the CpG context in DNA promoter sequences, where they seem to be involved in regulating gene expression.^[8] Lately, the synthesis and analysis of a tRNA segment of human mitochondrial tRNA^{Met} that carries 5-formylcytidine in the anticodon loop was reported.^[9]

To study the function of these new nucleosides, it is of utmost importance to develop methods that enable the synthesis of DNA strands that contain hmdC, fdC, and cadC. Recently, we reported the enzymatic incorporation of these nucleosides into DNA by means of their triphosphates.^[10] However, a sequence-specific incorporation of the nucleosides, particularly into CpG islands, which would be required for proteomics studies, is not possible with this method.

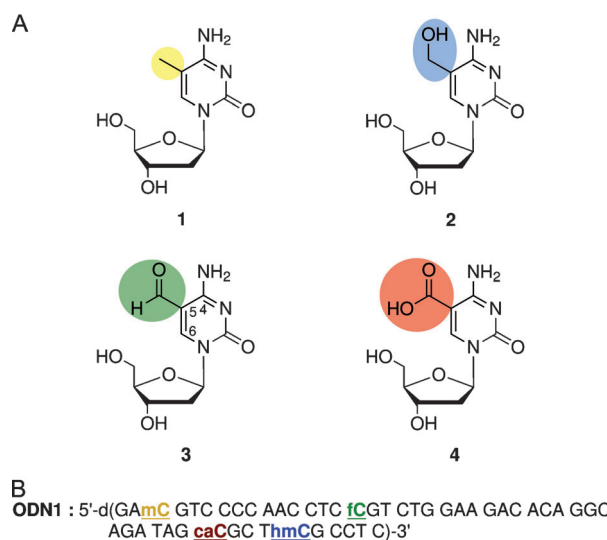


Figure 1. A) The recently found epigenetically relevant nucleosides 5-methyl- (**1**, mdC), 5-hydroxymethyl- (**2**, hmdC), 5-formyl- (**3**, fdC), and 5-carboxy-2'-deoxycytidine (**4**, cadC). B) Sequence of an OCT-4 promoter segment, in which cytidines in CpG contexts were substituted with the epigenetic nucleosides **1–4**.

The “Achilles’ heel” of the synthesis of DNA strands containing all new epigenetic nucleosides is the incorporation of fdC because of the high reactivity of the formyl group.^[11–13] The only currently available fdC building block is inserted as the 1,2-diol precursor, and periodate treatment is required to obtain the formyl moiety;^[14] this reagent is, for instance, incompatible with hmdC.^[15] Moreover, the periodate-based deprotection promotes side reactions that prevent the synthesis of DNA strands that contain multiple fdCs. Consequently, DNA strands that contain the four epigenetic nucleosides mdC, hmdC, fdC, and cadC at defined sites are synthetically not accessible.^[16]

Herein, we report the development of a new fdC phosphoramidite building block that can be inserted multiple times, or even consecutively, into DNA strands by using standard phosphoramidite chemistry. High coupling yields during solid-phase synthesis as well as mild deprotection conditions even enabled the preparation of a OCT-4 promoter sequence that contains all four epigenetic nucleosides (Figure 1 B, ODN1).

The electron-poor heterocycle of **3** is a major chemical obstacle for fdC incorporation, as this moiety promotes glycosidic bond cleavage even under the mildly acidic

[*] A. S. Schröder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder, Prof. Dr. T. Carell
Center for Integrated Protein Science
Department für Chemie, Ludwig-Maximilians-Universität
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: <http://www.carellgroup.de>

[**] We thank the Deutsche Forschungsgemeinschaft (SFB749 and SFB646) for financial support. Additional support from the Fonds der Chemischen Industrie (predoctoral fellowships for A.S.S., S.S., T.P.) and two excellence clusters (CiPS^M, EXC114) is acknowledged.

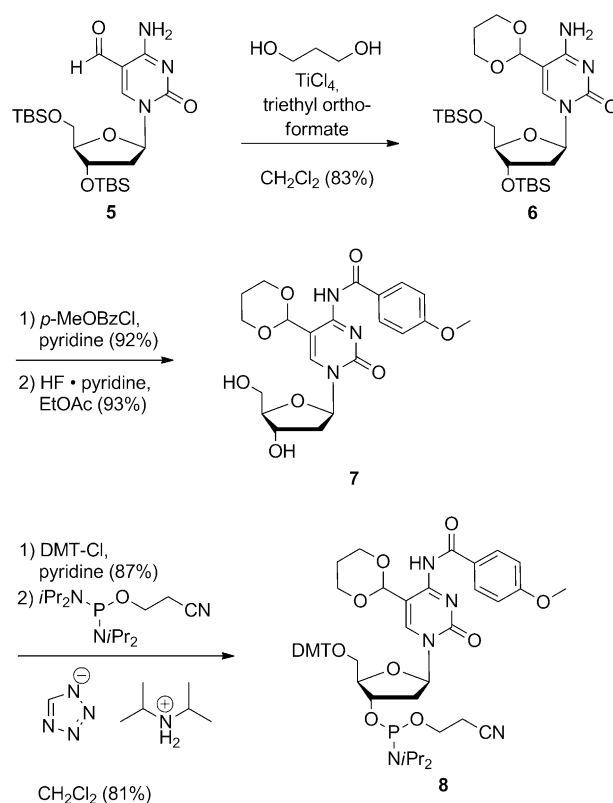
Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201308469>.

conditions that are commonly used for oligonucleotide synthesis. Furthermore, the formyl group of **3** is sensitive to oxidation, and, most importantly, 5-formylcytosine is readily attacked by nucleophiles, particularly at the C6 position (Figure 1A).^[16] To mask the reactivity of the aldehyde functionality, we examined various protecting groups (for details, see the Supporting Information). Finally, the best results were obtained when the aldehyde is converted into a 1,3-dioxane unit with propane-1,3-diol. This acetal turned out to be sensitive enough for later cleavage under mild conditions (80 % aqueous acetic acid, 20 °C); nevertheless, it is sufficiently stable to remain intact during solid-phase DNA synthesis. However, the temperature at which the deprotection is carried out at the end of the synthesis is a crucial parameter. Cleavage at 15 °C is too slow, whereas higher temperatures (≥ 25 °C) lead to a significant amount of glycosidic bond cleavage, mostly at fdC (for details, see the Supporting Information). Therefore, deprotection of the fdC acetal has to occur at approximately 20 °C. The amino group at the C4 position was protected with 4-methoxybenzoyl chloride, which is compatible with the acetal moiety. For fdC, we discovered that the commonly used benzoyl group is already efficiently cleaved during the solid-phase DNA synthesis, which gives rise to branched DNA products.

The synthesis of the fdC building block was achieved starting from *tert*-butyldimethylsilyl (TBS)-protected fdC **5**, which was synthesized from 5-iodo-dC according to literature procedures (Scheme 1).^[17] Notably, standard procedures for acetal protection did not provide the desired product **6**. Activation of the formyl group is strictly required, and TiCl_4 turned out to be the only Lewis acid that gave acceptable yields. Subsequent protection of the amino group at the C4 position with 4-methoxybenzoyl chloride and silyl deprotection furnished the fdC derivative **7**, which was converted into the fdC phosphoramidite building block **8** by standard procedures.^[18]

To examine the quality of the new fdC building block **8**, we prepared oligonucleotide **ODN2**, which contains five fdCs (Figure 2B). The solid-phase synthesis was performed using standard phosphoramidite conditions.^[19,20] For fdC, the coupling times were increased from 30 to 60 seconds to ensure maximum coupling yields. For deprotection of the standard nucleosides and cleavage of the oligonucleotide from the solid support, we first treated the solid-phase material with saturated aqueous NH_4OH solution (24 h, 25 °C). The HPL chromatogram depicted in Figure 2B shows the result for the crude product of the synthesis of the dimethoxytrityl (DMT)-protected oligonucleotide. In a second deprotection step, the oligonucleotide was treated with acetic acid for six hours at 20 °C, which cleaved the acetal and DMT protecting groups. The HPL chromatogram that was obtained for the unpurified DNA strand **ODN2**, which contains five fdCs, is shown in Figure 2C. As only one peak was observed, the synthesis had occurred in high yield.

To verify that five fdC nucleosides are indeed present in the DNA strand, we measured a MALDI-TOF spectrum (Figure 2D), which showed the expected exact mass. Further proof for correct incorporation was obtained by digestion of the synthesized, but still crude oligonucleotide with nucle-



Scheme 1. Synthesis of acetal-protected fdC phosphoramidite **8**.

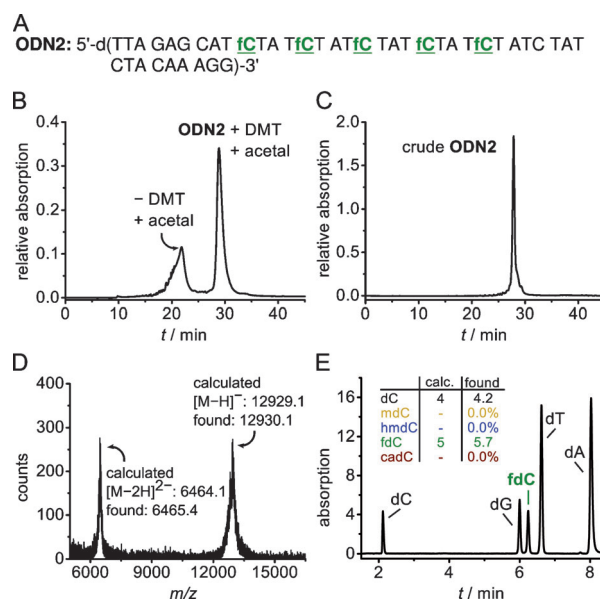


Figure 2. A) Sequence of the synthesized **ODN2** with fivefold incorporation of fdC using **8**. B) Reversed-phase HPL chromatogram of crude **ODN2** directly after basic cleavage from the resin (NH_4OH , 23 h, 25 °C, 0–80% buffer B in 45 min). C) Reversed-phase HPL chromatogram of crude **ODN2** directly after acidic cleavage of the DMT and acetal groups (80% aq. acetic acid, 6 h, 20 °C, 0–40% buffer B in 45 min). D) MALDI-TOF spectrum of fully deprotected, but crude **ODN2**. E) UHPL chromatogram and quantification data of unpurified digested **ODN2**.

ase S1 and snake-venom phosphodiesterase I to give the 5'-monophosphates. These were further hydrolyzed to the nucleosides by antarctic phosphatase. The obtained nucleoside mixture was analyzed by UHPLC-MS/MS (QQQ; Figure 2E). The obtained UV chromatogram was clean and showed signals only for the expected nucleosides. The fdC nucleoside was identified based on the retention time and its characteristic MS/MS fragmentation ($m/z = 256.1 \rightarrow 140.1$). Next, we quantified the amount of dC, fdC, and dT (Figure 2E, inset) by using synthetic isotopologues of the nucleosides as internal standards, which is a method that was recently described.^[16] We obtained the expected dC/fdC ratio of approximately 4:5, which is in full agreement with the sequence. The fact that these values are obtained with the crude synthesis product underpins the superior properties of the new fdC phosphoramidite. Most importantly, we did not detect any α -fdC nucleosides, which indicates that β - to α -anomerization was fully suppressed with the new building block. We also investigated whether hmdC and cadC are present during DNA synthesis; these are the reduction and oxidation products of fdC, which might be formed during the iodine oxidation step. As expected, none of these compounds were detected (Figure 2E, inset). In summary, the fdC phosphoramidite **8** enabled the synthesis of fdC-containing oligonucleotides. With this building block in hand, we have established the basis for the synthesis of oligonucleotides that contain all four epigenetic nucleosides.

Therefore, we then synthesized oligonucleotide **ODN1** with the sequence of a segment of the OCT-4 promoter (*mus musculus*, chromosome 17, 35505895-35505943; Figure 3B). All cytidines in the respective CpG contexts were substituted with the epigenetic nucleosides **1–4**. The best results were obtained with the cyclic-carbamate-protected hmdC phosphoramidite (for detailed conditions of the oligonucleotide synthesis, see the Supporting Information).^[17] This carbamate building block and the ester-protected cadC unit forced us to use NaOH (0.4 M in methanol/water (4:1)) instead of NH_4OH in the first deprotection step to avoid the formation of amide moieties (Figure 3A).^[12,17] After treatment of the synthesized **ODN1** with NaOH, the oligonucleotide was precipitated with ethanol. We next performed the second deprotection step with acetic acid. Figure 3C shows the HPL chromatogram of the crude material after the second deprotection step. It is clearly visible that the process yielded **ODN1** as the major product, despite the fact that **ODN1** is a 49 mer oligonucleotide that contains four modified nucleosides (for detailed data of crude **ODN1**, see the Supporting Information). Purification by HPLC after the first deprotection step and subsequent acidic treatment as described above gave pure **ODN1** in 31 % (Figure 3C, inset; see also the Supporting Information). The oligonucleotide was fully digested, and the product of hydrolysis was analyzed by UHPLC-MS/MS, which corroborated the presence of all incorporated nucleosides. Isotope-based quantification confirmed the amounts of the epigenetic nucleosides relative to dC to be 1.2:1.2:0.9:1.0:15.0, which is in full agreement with the sequence (Figure 3B,D).

In summary, the new fdC building block **8** enabled the synthesis of fdC-containing DNA strands in unprecedented yield and quality, which will facilitate biological studies of

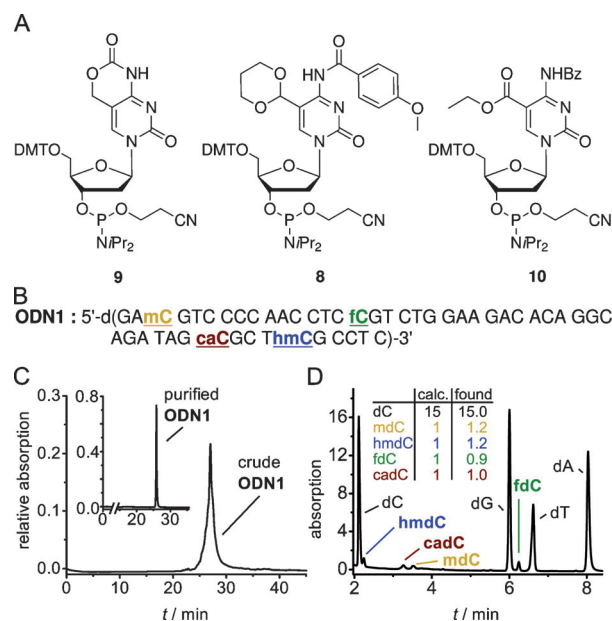


Figure 3. A) Used phosphoramidites. B) Sequence of the synthesized OCT-4 promoter segment, in which cytidines in CpG contexts were substituted with the epigenetic nucleosides **1–4**. C) Reversed-phase HPL chromatogram of crude and purified (inset) **ODN1** after complete deprotection (0–40 % buffer B in 45 min). D) UHPL chromatogram and quantification data of digested **ODN1**. Bz = benzoyl.

fdC. In combination with phosphoramidite building blocks for mdC, hmdC, and cadC, it is now possible to synthesize oligonucleotides that contain all four epigenetically important cytidine nucleosides **1–4**. This will pave the way for a more detailed analysis of how these nucleosides influence biological processes and stem-cell development.^[8]

Received: September 28, 2013

Published online: November 26, 2013

Keywords: 5-formylcytosine · DNA methylation · epigenetics · oligonucleotides · solid-phase synthesis

- [1] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.
- [2] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
- [3] M. Tahilian, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
- [4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
- [5] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
- [6] C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583.
- [7] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [8] C. X. Song, C. He, *Trends Biochem. Sci.* **2013**, *38*, 480–484.
- [9] E. M. Gustilo, F. A. P. Vendeix, R. Kaiser, M. O. Delaney, W. D. Graham, V. A. Moye, W. A. Cantara, P. F. Agris, A. Deiters, *Nucleic Acids Res.* **2008**, *36*, 6548–6557.

- [10] B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366–369.
 - [11] S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nucleic Acids Res.* **1997**, *25*, 553–559.
 - [12] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782–13788.
 - [13] Q. Dai, C. He, *Org. Lett.* **2011**, *13*, 3446–3449.
 - [14] N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456–2463.
 - [15] T. M. A. Shaikh, L. Emmanuvel, A. Sudalai, *J. Org. Chem.* **2006**, *71*, 5043–5046.
 - [16] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
 - [17] M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671–5673.
 - [18] W. Bannwarth, A. Trzeciak, *Helv. Chim. Acta* **1987**, *70*, 175–186.
 - [19] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859–1862.
 - [20] N. D. Sinha, J. Biernat, J. McManus, H. Köster, *Nucleic Acids Res.* **1984**, *12*, 4539–4557.
-

3.5 Chaperoning epigenetics: FKBP51 decreases the activity of DNMT1 and mediates epigenetic effects of the antidepressant paroxetine

Nils C. Gassen⁺, Gabriel R. Fries⁺, Anthony S. Zannas, Jakob Hartmann, Jürgen Zschocke, Kathrin Hafner, Tania Carrillo-Roa, Jessica Steinbacher, S. Nicole Preißinger, Lianne Hoeijmakers, Matthias Knop, Frank Weber, Stefan Kloiber, Susanne Lucae, George P. Chrousos, Thomas Carell, Marcus Ising, Elisabeth B. Binder, Mathias V. Schmidt, Joëlle Rüegg, Theo Rein (⁺ geteilte Erstautorenschaft)

Prolog

Epigenetische Prozesse wie die DNA Methylierung und molekulare Chaperone wie das FK506 Bindeprotein 51 (FKBP51) sind unabhängig voneinander an stressbedingten mental-psychischen Erkrankungen und der Wirkung von Antidepressiva beteiligt.

FKBP51 steht im Zusammenhang mit der Cyclin abhängigen Kinase 5 (CDK5), welche eine der Kinasen ist, die die DNA Methyltransferase 1 (Dnmt1) phosphoryliert und damit aktiviert. In diesem Manuskript wird in Zellen von Mäusen und in gesunden sowie in unter Depression leidenden Menschen nach einer Verbindung zwischen FKBP51 und Dnmt1 gesucht. Es wurde herausgefunden, dass FKBP51 mit seinem Homologen FKBP52 um eine Verbindung mit CDK5 konkurriert. In *human embryonic kidney* (HEK) 293 Zellen wird durch die Expression von FKBP51 FKBP52 von CDK5 verdrängt, die Wechselwirkung zwischen CDK5 und Dnmt1 geschwächt, die Phosphorylierung und die enzymatische Aktivität reduziert und die globale DNA Methylierung verringert. In embryonalen Fibroblasen und primären Astrozyten der Maus vermittelt FKBP51 zahlreiche Effekte von Paroxetin. Es reduziert die Protein-Protein Wechselwirkung von Dnmt1 mit CDK5 und FKBP52, verringert die Phosphorylierung von Dnmt1, setzt die Methylierungslevel herab und verstärkt die Expression des *gene encoding brain-derived neurotrophic factor* (Bdnf). In menschlichen peripheren Blutzellen korreliert die FKBP5 Expression negativ sowohl mit der globalen Methylierung als auch mit der Methylierung von Bdnf. Periphere Blutzellen aus depressiven Patienten, die *ex vivo* mit Paroxetin behandelt wurden machten deutlich, dass die Menge an BDNF positiv und

phosphoryliertes Dnmt1 negativ mit FKBP51 in den Zellen korreliert.

Autorenbeitrag

Der Verdau der genomischen DNA, die Entwicklung einer geeigneten Messmethode sowie die massenspektrometrische Quantifizierung und die Auswertung der Daten wurde von mir vorgenommen. Für eine weitere Auflistung des Beitrags der übrigen Autoren siehe Manuskript.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags.

Chaperoning epigenetics: FKBP51 decreases the activity of DNMT1 and mediates epigenetic effects of the antidepressant paroxetine

Nils C. Gassen,^{1*†} Gabriel R. Fries,^{1,2†} Anthony S. Zannas,^{1,3} Jakob Hartmann,⁴ Jürgen Zschocke,¹ Kathrin Hafner,¹ Tania Carrillo-Roa,¹ Jessica Steinbacher,⁵ S. Nicole Preißinger,¹ Lianne Hoeijmakers,⁴ Matthias Knop,⁶ Frank Weber,⁶ Stefan Kloiber,⁶ Susanne Lucae,⁶ George P. Chrousos,⁷ Thomas Carell,⁵ Marcus Ising,⁶ Elisabeth B. Binder,^{1,8} Mathias V. Schmidt,⁴ Joëlle Rüegg,⁹ Theo Rein^{1*}

Epigenetic processes, such as DNA methylation, and molecular chaperones, including FK506-binding protein 51 (FKBP51), are independently implicated in stress-related mental disorders and antidepressant drug action. FKBP51 associates with cyclin-dependent kinase 5 (CDK5), which is one of several kinases that phosphorylates and activates DNA methyltransferase 1 (DNMT1). We searched for a functional link between FKBP51 (encoded by *FKBP5*) and DNMT1 in cells from mice and humans, including those from depressed patients, and found that FKBP51 competed with its close homolog FKBP52 for association with CDK5. In human embryonic kidney (HEK) 293 cells, expression of FKBP51 displaced FKBP52 from CDK5, decreased the interaction of CDK5 with DNMT1, reduced the phosphorylation and enzymatic activity of DNMT1, and diminished global DNA methylation. In mouse embryonic fibroblasts and primary mouse astrocytes, FKBP51 mediated several effects of paroxetine, namely, decreased the protein-protein interactions of DNMT1 with CDK5 and FKBP52, reduced phosphorylation of DNMT1, and decreased the methylation and increased the expression of the gene encoding brain-derived neurotrophic factor (*Bdnf*). In human peripheral blood cells, *FKBP5* expression inversely correlated with both global and *BDNF* methylation. Peripheral blood cells isolated from depressed patients that were then treated *ex vivo* with paroxetine revealed that the abundance of BDNF positively correlated and phosphorylated DNMT1 inversely correlated with that of FKBP51 in cells and with clinical treatment success in patients, supporting the relevance of this FKBP51-directed pathway that prevents epigenetic suppression of gene expression.

INTRODUCTION

Environmental factors and their influence on gene expression are recognized as key players in several psychiatric diseases, including major depressive disorder (MDD) (1–3). Although yet unclear, it is proposed that these environmental effects are mediated by epigenetic changes, such as methylation and demethylation of DNA and posttranslational modifications of histones. Such epigenetic marks control the accessibility of transcriptional machinery to the DNA and are responsive to both environmental stressors and to genetic variations. Whereas histone modifications occur at various sites and involve the action of various enzymes, DNA methylation

is largely confined to CpG dinucleotides in the mammalian genome and is executed by DNA methyltransferases (DNMTs) (4), which comprise a family of proteins with three subtypes that exhibit different specificities and functions: DNMT1, DNMT3a, and DNMT3b (5).

Epigenetic mechanisms are considered not only crucial in shaping the phenotype of complex psychiatric disorders but also important for the response to certain medications (6–9). For instance, some antidepressants can reduce DNMT1 activity (10), the major DNMT ensuring the maintenance DNA methylation during S phase that has also been implicated in *de novo* methylation (11, 12); this effect of antidepressants appears to be partly due to the reduction of the amounts of the histone methyltransferase G9a (13). Furthermore, DNMT inhibitors applied either systemically or locally in the hippocampus induce antidepressant-like effects in mice, which are accompanied by increased expression of the gene encoding brain-derived neurotrophic factor (BDNF) (14). Epigenetic regulation of BDNF is implicated in the development and treatment of psychiatric diseases in several studies (9, 15–17).

The activity of DNMT1 is modulated by several interacting proteins and by posttranslational modifications (18, 19), including phosphorylation (20–22). Among the phosphorylated sites of DNMT1 is Ser¹⁵⁴, which is targeted by cyclin-dependent kinases (CDKs), such as CDK5, and increases DNMT1 activity (23). Thus, several mechanisms might be considered for the antidepressant-induced effects on DNMT1.

Psychological stress and trauma are consistently associated with MDD (1). The glucocorticoid receptor (GR) is integral to the stress response and is controlled by a complex of chaperones and cochaperones (24). The

¹Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, 80804 Munich, Germany. ²Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center, Houston, TX 77054, USA. ³Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC 27710, USA. ⁴Department of Stress Neurobiology and Neurogenetics, Max Planck Institute of Psychiatry, 80804 Munich, Germany. ⁵Department of Organic Chemistry, Faculty of Chemistry and Pharmacy, Ludwig Maximilians University, 81377 Munich, Germany. ⁶Department of Clinical Research, Max Planck Institute of Psychiatry, 80804 Munich, Germany. ⁷First Department of Pediatrics, University of Athens Medical School, Athens 11527, Greece. ⁸Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA 30329, USA. ⁹Department of Clinical Neuroscience, Center for Molecular Medicine, Karolinska Institutet, 171 76 Stockholm, Sweden and Swedish Toxicology Science Research Center, Swetox, 151 36 Södertälje, Sweden.

*Corresponding author. E-mail: ncgassen@psych.mpg.de (N.C.G.); theorein@psych.mpg.de (T.R.)

†First co-authors.

cochaperones FKBP51 and FKBP52 are major determinants of GR activity by competing with each other for access to the GR-chaperone hetero-complex and by exerting opposing effects on GR function with FKBP52-promoting and FKBP51-inhibiting GR activity (25–27). Conversely, FKBP51 abundance increases after GR activation (28–30). Notably, FKBP51 has been linked genetically to the response to antidepressants, and enhanced abundance of FKBP51 is associated with improved treatment response (31–36). In addition, FKBP51 and FKBP52 interact with several proteins including CDK5 (36, 37). Therefore, we aimed to explore potential mechanisms by which the stress-related cochaperones FKBP51 and FKBP52 modulate DNMT1 phosphorylation and activity and whether this might contribute to the clinical response to antidepressant treatment in patients with MDD.

RESULTS

FKBP51 and FKBP52 were differentially associated with DNMT1 and modulate its phosphorylation and activity

Previously, we found that CDK5 formed a protein complex with FKBP51 or FKBP52 (36). Because CDK5 has been reported to regulate DNMT1 by phosphorylation at Ser¹⁵⁴ (23), we investigated whether FKBP51 and FKBP52 modulate CDK5's action on DNMT1. We initially tested for associations between CDK5 and DNMT1 with ectopic FLAG-tagged FKBP51 or FKBP52 in human embryonic kidney (HEK) 293 cells by co-immunoprecipitation. Both FKBP51 and FKBP52 were associated with CDK5 (Fig. 1A), consistent with our previous results (36); however, only FKBP52 was associated with DNMT1 (Fig. 1B). Coexpressing both FKBP51 and FKBP52 revealed that they compete with each other for binding to CDK5 (Fig. 1C). Notably, coexpression of FKBP51 also reduced the interaction between FKBP52 and DNMT1 (Fig. 1C).

Because FKBP51 and FKBP52 are cochaperones of heat shock protein 90 (Hsp90) (25) and Hsp90 interacts with both CDK5 and DNMT1 (38, 39), we explored the possibility that Hsp90 function might be important for complex assembly. We exposed HEK293 cells to increasing concentrations of 17-AAG, an Hsp90 inhibitor and geldanamycin derivative, immunoprecipitated CDK5, and probed for the interaction with Hsp90 and DNMT1. There was no significant effect

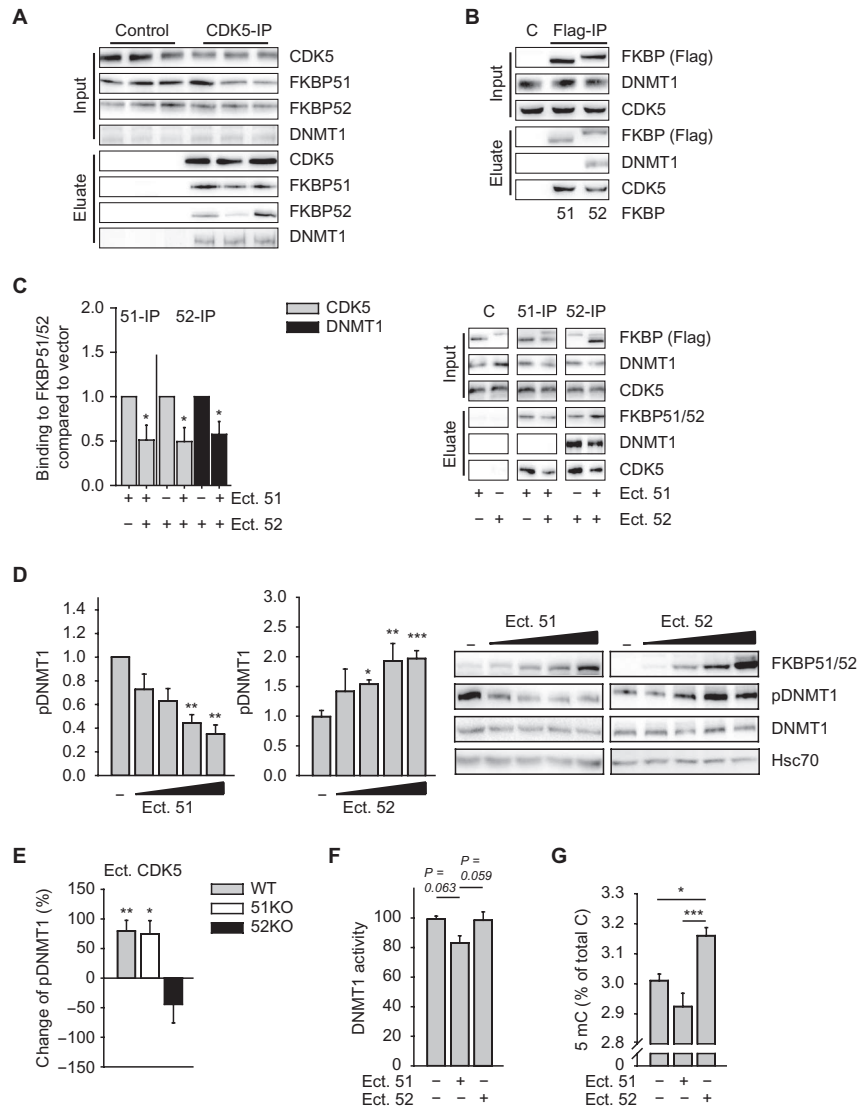


Fig. 1. FKBP51 and FKBP52 differentially affect CDK5-dependent phosphorylation and activity of DNMT1. (A) Immunoprecipitation (IP) for CDK5 followed by Western blotting in lysates from HEK293 cells transfected with FKBP51 or FKBP52. Control: without primary antibody [immunoglobulin G (IgG) control in fig. S1A]. (B and C) Immunoprecipitation for the FLAG tag (B) or FKBP51 or FKBP52 (C) followed by Western blotting in lysates from HEK293 cells transfected with FLAG-tagged FKBP51, FKBP52, or both, as indicated. C, control [vector-transfected cells (B) or without primary antibody (C) (IgG control in fig. S1A)]. Blots are representative of three independent experiments. Ect., ectopic expression. Graph (C) displays the association of FKBP51 (51-IP) or FKBP52 (52-IP) to CDK5 or DNMT1 in the dual transfected samples relative to the singly transfected samples. Data are means + SEM of three independent experiments. (D) Western blotting for DNMT1 and phosphorylated (p) DNMT1 (at Ser¹⁵⁴) in primary rat astrocytes transfected with FKBP51 or FKBP52. Amount of phosphorylated DNMT1 was calculated relative to that of total DNMT1; this was then calculated relative to that in the control vector sample (–), arbitrarily set to 1. Data are mean + SEM of three to six independent experiments. (E) Quantification of Western blotting analysis of the phosphorylation of DNMT1 (corrected for total DNMT1) in wild-type (WT), *Fkbp51* knockout (51KO), or *Fkbp52* knockout (52KO) MEFs transfected with CDK5 relative to vector-transfected cells. Data are means + SEM of three independent experiments. Representative blot is shown in fig. S2B; full blots for (D) and (E) are shown in data file S1. (F and G) DNMT1 enzymatic activity in total nuclear lysates from primary astrocytes (F) and cytosine methylation (percent of total cytosine) assessed by mass spectrometry in isolated total DNA from primary rat astrocytes transfected with FKBP51 or FKBP52 or a control vector. 5mC, 5-methylcytosine. Data are means + SEM of three independent experiments each performed with two (F) or three (G) technical replicates. **P* < 0.05, ***P* < 0.01. Statistical details in tables S2 and S3.

of Hsp90 inhibition on complex assembly of CDK5 and DNMT1 (fig. S1, A to D, and data file S1).

In addition to DNMT1, which typically maintains DNA methylation, DNMT3a and DNMT3b also shape the methylome as de novo methyltransferases (5). Coimmunoprecipitation revealed an association between FKBP52 and DNMT3a but not DNMT3b, whereas FKBP51 associated with none (fig. S1, E and F).

Considering the reported effects of CDK5 on DNMT1, we analyzed the phosphorylation of DNMT1 at Ser¹⁵⁴ upon overexpression of FKBP51 or FKBP52 in HEK293 cells or primary rat astrocytes. Overexpressing FKBP51 significantly lowered the phosphorylation of DNMT1, whereas overexpressing FKBP52 increased it (Fig. 1D, data file S1, and fig. S2A). Because phosphorylation of Ser¹⁵⁴ is important for DNMT1 activity (23), this suggests that FKBP52 promotes whereas FKBP51 inhibits DNMT1 activity. To check for the relevance of the FKBP5s on CDK5-induced DNMT1 Ser¹⁵⁴ phosphorylation, we overexpressed CDK5 in wild-type, 51KO, and 52KO mouse embryonic fibroblasts (MEFs) and monitored DNMT1 phosphorylation. CDK5 overexpression did not change the phosphorylation of DNMT1 in 52KO MEFs (as opposed to wild-type cells), whereas a marked increase in phosphorylation of DNMT1 was evident in wild-type and 51KO MEFs (Fig. 1E, data file S1, and fig. S2B).

To determine whether the FKBP51/52-dependent changes in phosphorylation of DNMT1 go along with changes in DNMT1 activity, we transfected primary rat astrocytes with plasmids expressing FKBP51 or FKBP52 and determined the enzymatic DNA methylation activity in total cell lysates. There was a trend toward lower methylase activity in FKBP51-transfected cells than in FKBP52-transfected cells (Fig. 1F). Analysis of the content of total methylated cytosines in genomic DNA revealed significantly lower DNA methylation in FKBP51-transfected than in FKBP52-transfected cells (Fig. 1G). There was no difference in hydroxymethylation (fig. S2C).

To further shed light on the possible mechanism involved in the FKBP-directed regulation of DNMT1 activity by CDK5, we analyzed the association of the CDK5 regulatory proteins p25, p35, and p39 (40) in the human neuroblastoma cell line SK-N-MC. Although all three proteins were detected in protein lysates, only the CDK5 activator p35 associated with DNMT1 (Fig. 2A), indicating that active CDK5 is recruited to DNMT1. Ectopic expression of FKBP52 enhanced the association of p35 with DNMT1, whereas ectopic expression of FKBP51 did not (Fig. 2, A and B, and data file S2). There was no significant change in the association of DNMT1 with CDK5 (Fig. 2C).

FKBP51 has been suggested to facilitate the dephosphorylation of tau, depending on its peptidylprolyl isomerase (PPIase) activity (41). To test the relevance of PPIase activity in the regulation of DNMT1 phosphorylation, we expressed increasing amounts of a PPIase-deficient mutant of FKBP51 (25, 42) in HEK293 cells. PPIase-deficient FKBP51 exerted similar effects on phosphorylation of DNMT1 as the nonmutated FKBP51 (Fig. 2D and data file S2).

Modulation of DNMT1 phosphorylation and activity by paroxetine depend on FKBP51

On the basis of previous evidence of an inhibitory effect of paroxetine on DNMT1 (13), we sought to assess whether the antidepressant action on DNMT1 involves DNMT1 phosphorylation or its association with CDK5, FKBP51, and FKBP52. First, we tested whether paroxetine affects the association between the FKBP5s and CDK5 in HEK293 cells. We found that treatment with paroxetine increased the association between FKBP51 and CDK5 and reduced the interaction between FKBP52 and CDK5 (Fig. 3A, fig. S3, and data file S1). There was no significant effect on the interaction between FKBP52 and DNMT1.

We next checked for the effects of paroxetine on phosphorylation of DNMT1 in primary astrocytes and found that it decreased phosphorylation of

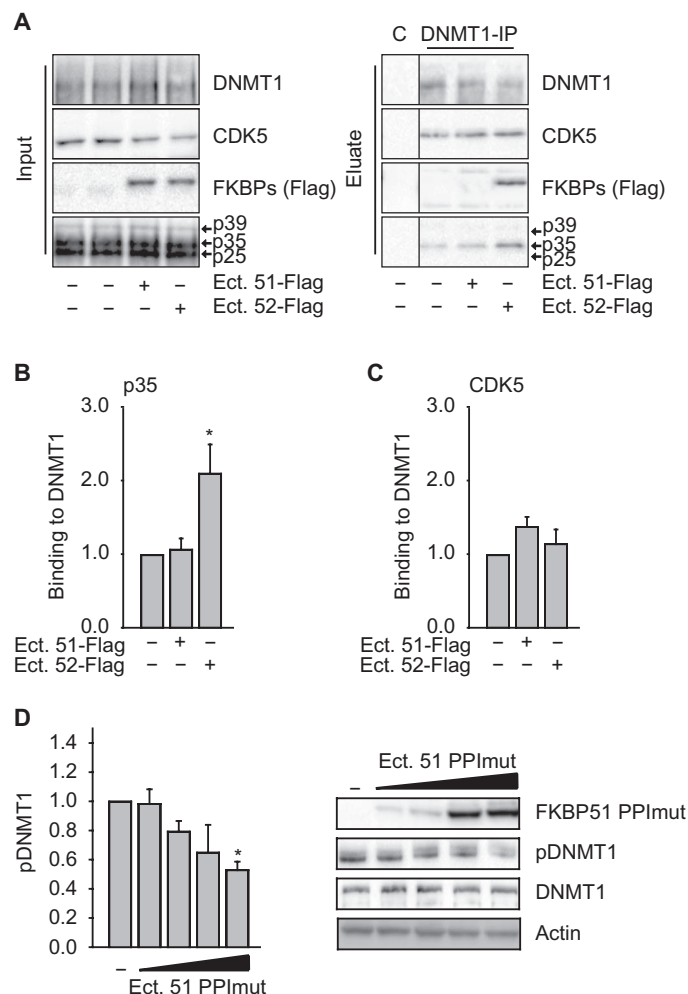


Fig. 2. FKBP51 does not change the interaction of DNMT1 with the CDK5 activator protein p35 and acts independently of its PPIase activity on the phosphorylation of DNMT1. (A to C) Immunoprecipitation for DNMT1 followed by Western blotting as indicated (A) and quantification of DNMT1-p35 (B) and DNMT1-CDK5 (C) interaction in SK-N-MC cells transfected with FLAG-tagged FKBP51 or FKBP52 or vector control (indicated by “-” in both lanes). IgG control in fig. S1A. (D) Western blotting as indicated in HEK293 cells transfected with PPIase-deficient mutant FKBP51 (F67D/D68V: 51 PPImut). Phosphorylated DNMT1 was calculated against total DNMT1, relative to that in the control vector sample (–), arbitrarily set to 1. Data are means \pm SEM from three independent experiments. * $P < 0.05$. Statistical details in table S2. Full blots for (B) and (D) are shown in data file S2.

ation of DNMT1 in a dose-dependent manner (Fig. 3B). Furthermore, we found evidence that the antidepressant-induced reduction in phosphorylation of DNMT1 requires the presence of FKBP51 because paroxetine decreased phosphorylation of DNMT1 in primary astrocytes from wild-type mice but not in primary astrocytes from 51KO mice (Fig. 3C and data file S3). Similar results were obtained in wild-type and 51KO MEFs (Fig. 3D, fig. S4A, and data file S3). Likewise, paroxetine reduced DNMT1 enzymatic activity in wild-type but not in 51KO MEFs (Fig. 3E and data file S3). To test whether these results can be translated to the action of paroxetine in an established animal model, we determined phosphorylation of DNMT1 in mice that were acutely or chronically treated with

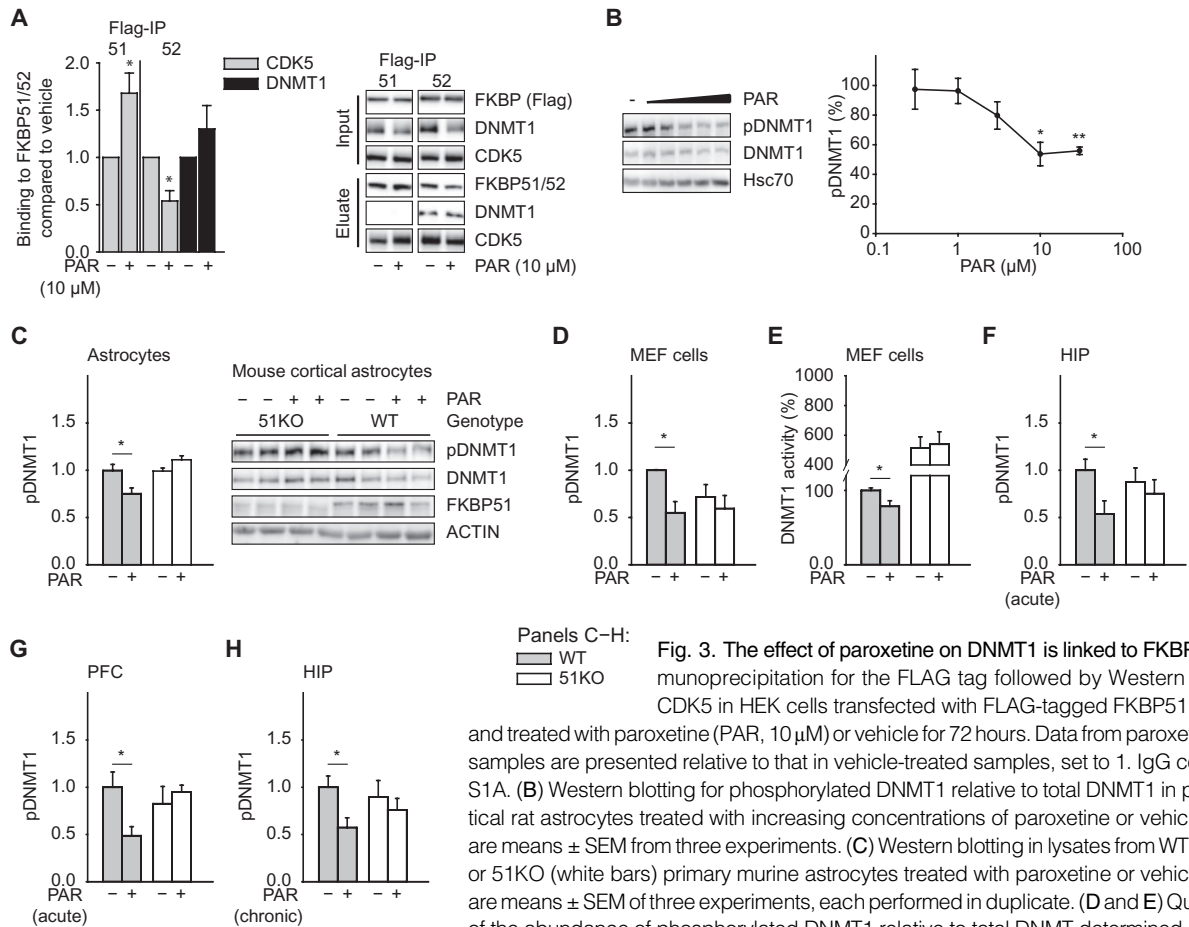


Fig. 3. The effect of paroxetine on DNMT1 is linked to FKBP51. (A) Immunoprecipitation for the FLAG tag followed by Western blotting for CDK5 in HEK cells transfected with FLAG-tagged FKBP51 or FKBP52 and treated with paroxetine (PAR, 10 μ M) or vehicle for 72 hours. Data from paroxetine-treated samples are presented relative to that in vehicle-treated samples, set to 1. IgG control in fig. S1A. (B) Western blotting for phosphorylated DNMT1 relative to total DNMT1 in primary cortical rat astrocytes treated with increasing concentrations of paroxetine or vehicle (–). Data are means \pm SEM from three experiments. (C) Western blotting in lysates from WT (gray bars) or 51KO (white bars) primary murine astrocytes treated with paroxetine or vehicle (–). Data are means \pm SEM of three experiments, each performed in duplicate. (D and E) Quantification of the abundance of phosphorylated DNMT1 relative to total DNMT1 determined by Western

blotting (D) and enzymatic DNMT1 activity (E) in WT (gray bars) or 51KO (white bars) MEFs treated with paroxetine or vehicle (–). Data are means \pm SEM of three (D) or four (E) independent experiments performed in technical duplicates. (F to H) Abundance of phosphorylated relative to total DNMT1 in the prefrontal cortex (PFC) and hippocampus (HIP) of 51KO mice (white bars) and WT mice (gray bars) treated with paroxetine or vehicle in an acute (45 min) (F and G) or chronic (21 days) (H) regimen (33). Data are means \pm SEM of 8 to 10 animals. * P < 0.05, ** P < 0.01. Statistical details are in tables S2 and S3. Full blots for (C) and (F) to (H) are shown in data file S3.

paroxetine (45-min or 21-days treatment duration, respectively). Tissue was available from previous experiments that documented that the behavioral effect of paroxetine was absent in mice lacking FKBP51 (33). Acute and chronic treatment with paroxetine led to reduced phosphorylation of DNMT1, both in the hippocampus and in the prefrontal cortex, only in the presence of FKBP51 (Fig. 3, F to H; fig. S4, B to D; and data file S3).

We also investigated the effects of the tricyclic antidepressant amitriptyline on phosphorylation of DNMT1. In MEFs, there was no significant effect of amitriptyline (fig. S5A). In wild-type mice previously treated with amitriptyline (33), a reduction in phosphorylation of DNMT1 was observed in the prefrontal cortex but not in the hippocampus (fig. S5, B and C). This effect of amitriptyline was abolished in 51KO mice (fig. S5C). To also test another neuropharmacologically active substance, we analyzed protein extracts from mice that had been treated with the neuroleptic haloperidol (33). Haloperidol displayed no significant effect on phosphorylation of DNMT1 in the hippocampus and slightly enhanced phosphorylation of DNMT1 in the prefrontal cortex (fig. S5, D to G), contrasting the effects of the antidepressants paroxetine and amitriptyline.

FKBP51- and paroxetine-modulated DNMT1 activity correlates with *Bdnf* promoter methylation and expression in mice

In addition to the effects of FKBP51 and paroxetine on global DNA methylation through their action on DNMT1, we analyzed possible local consequences at *Bdnf*, the gene encoding BDNF, as a locus relevant to neuropsychiatry. *Bdnf* features different epigenetically controlled promoters that give rise to multiple isoforms at the mRNA level with different untranslated exons at the 5' end spliced to a common protein-coding exon at the 3' end (43, 44). To date, the *Bdnf* promoter of exon IV (called promoter IV) is one of the best characterized (9, 45, 46). The expression of promoter IV is representative of *Bdnf* expression in the brain and also reflects the changes induced by neuronal activity and antidepressants (16, 17, 47, 48). Therefore, we analyzed the methylation of promoter IV in the brain of wild-type and 51KO mice that had been treated with paroxetine for 45 min (33). In these animals, paroxetine induced the demethylation of *Bdnf* at promoter IV in the prefrontal cortex of wild-type but not 51KO mice (Fig. 4A and table S1). The extent of demethylation varied between 10 and 40 percentage points, depending on the CpG site.

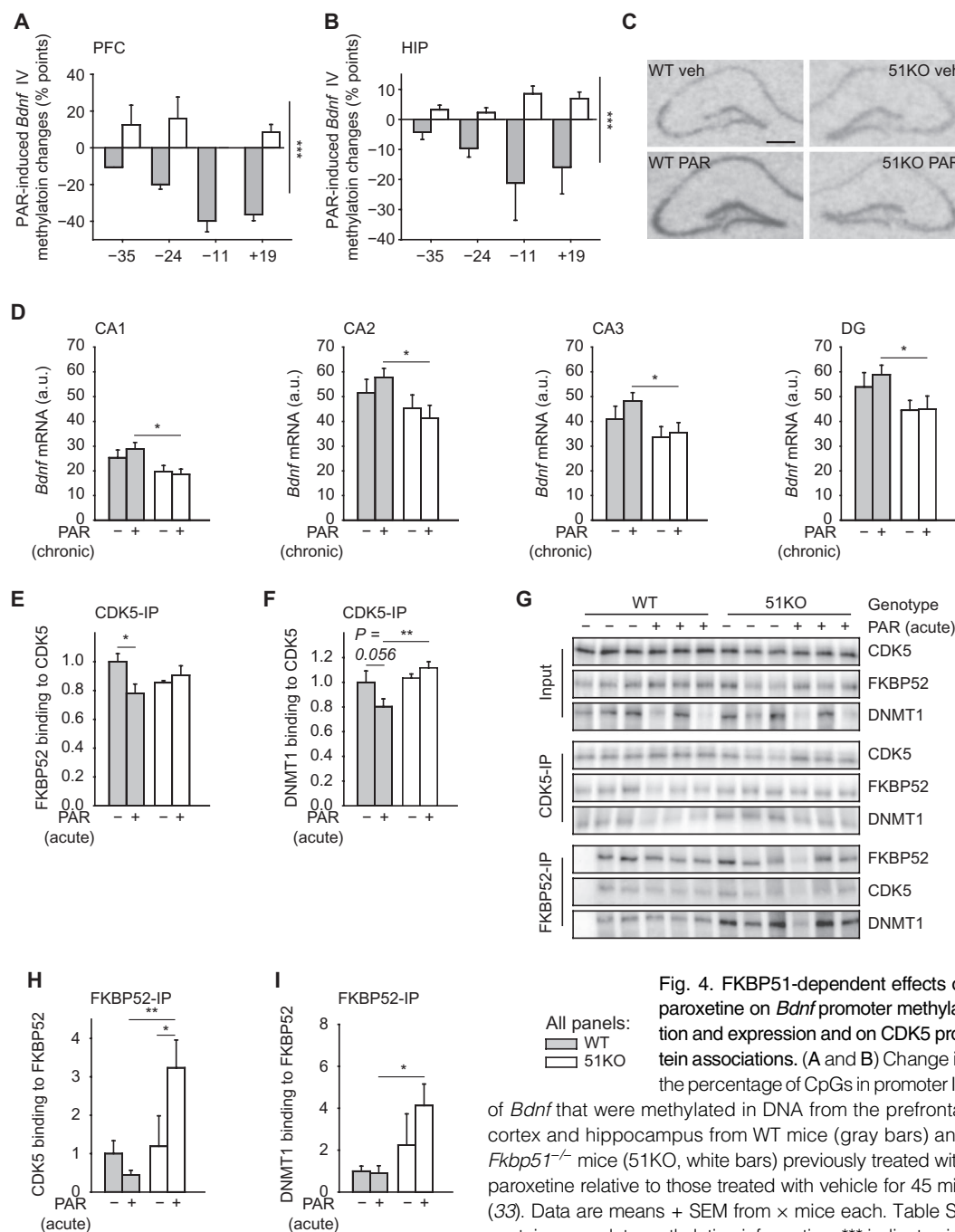


Fig. 4. FKBP51-dependent effects of paroxetine on *Bdnf* promoter methylation and expression and on CDK5 protein associations. (A and B) Change in the percentage of CpGs in promoter IV of *Bdnf* that were methylated in DNA from the prefrontal cortex and hippocampus from WT mice (gray bars) and *Fkbp51*^{-/-} mice (51KO, white bars) previously treated with paroxetine relative to those treated with vehicle for 45 min (33). Data are means + SEM from \times mice each. Table S1 contains complete methylation information. *** indicate significant overall genotype effects. (C and D) *Bdnf* expression assessed by in situ hybridization shown in the dorsal hippocampus (C) and quantified in hippocampal sections from WT and 51KO mice treated with paroxetine for 21 days (chronic) or vehicle (-) (33). Scale bar, 500 μ m. a.u., arbitrary units; DG, dentate gyrus. (E to I) Protein binding to CDK5 (E and F) or FKBP52 (H and I) in brain extracts from WT and 51KO mice that had been treated with paroxetine for 45 min (33). Western blot representing three animals for each condition and treatment shown in (G). IgG control in fig. S1A. Data are means + SEM from 9 to 10 animals for each condition or treatment. Data from WT mice treated with vehicle were set to 1. * P < 0.05, ** P < 0.01, *** P < 0.001. Statistical details in tables S2 and S3.

In 51KO mice, there was some increase in DNA methylation upon paroxetine treatment, but this change was not statistically significant (Fig. 4A and table S1). There was no significant genotype effect in the absence of

paroxetine (table S1). Less pronounced effects were observed in the hippocampus, but there was still a significant genotype effect (Fig. 4B and table S1). In addition, we analyzed the promoter of exon I, but no significant changes in DNA methylation were observed (table S1). We also analyzed the effect of amitriptyline on the methylation of *Bdnf* at the exon IV promoter in the hippocampus and in the prefrontal cortex of wild-type and 51KO mice, but there was no significant change in methylation (fig. S6).

To assess changes in the transcription of *Bdnf*, we performed in situ hybridization in brain slices from wild-type and 51KO mice that had previously been treated with paroxetine or vehicle for 21 days (33). In all subregions of the hippocampal formation, we observed a similar pattern: the amount of *Bdnf* mRNA was significantly higher in wild-type mice treated with paroxetine than it was in 51KO mice treated with paroxetine (Fig. 4, C and D).

These data are in line with the hypothesis that FKBP51-dependent reduction of DNMT1 activity by paroxetine, also indicated by the decrease of phosphorylation of DNMT1 (Fig. 3, F to H, and fig. S4, B to D), leads to demethylation and activation of the *Bdnf* gene. To evaluate whether FKBP51 mediates the effects of paroxetine on CDK5 and FKBP52 complexes in mice, we performed CDK5 and FKBP52 immunoprecipitation in brain extracts from wild-type and 51KO mice that had been treated with paroxetine for 45 min (33). Paroxetine changed the association of CDK5 with FKBP52 and with DNMT1 in wild-type but not 51KO mice (Fig. 4, E to G). This is consistent with the inhibitory effect of paroxetine on FKBP52-CDK5 association we observed in cells (Fig. 3A). When we precipitated FKBP52, the association

with CDK5 or DNMT1 in the presence of paroxetine was greater in 51KO than in wild-type mice (Fig. 4, G to I). Thus, FKBP51 mediates the impact of paroxetine on protein associations.

Cellular and clinical treatment responses support the physiological relevance of FKBP51-dependent effects of paroxetine on DNMT1 and BDNF

To translate the findings obtained from cellular and animal studies to humans, we made use of the whole-blood DNA methylation and RNA data available from the Grady Trauma Project, an ongoing study examining the role of genetic and environmental factors that predict stress responses (49, 50). The expression of *FKBP5* (the gene encoding FKBP51) was negatively correlated with the average global DNA methylation, as measured by 450K arrays ($r = -0.127$, $P = 0.015$, $n = 365$), and with the average amount of *BDNF* methylation found over the entire *BDNF* locus (total of 79 CpGs), as well as with the average methylation of CpGs located near the transcription start sites of exon I (33 CpGs) and exon IV (31 CpGs) of the gene (Fig. 5, A and B, and fig. S7).

To test protein correlations, we first analyzed the phosphorylation of DNMT1 and FKBP51 in peripheral blood mononuclear cells (PBMCs) from healthy individuals. We observed that the amounts of FKBP51 and phosphorylated DNMT1 were inversely correlated, in line with a negative effect of FKBP51 on the phosphorylation of DNMT1 (Fig. 5C). We also checked for the effects of paroxetine on the same cells by treating PBMCs ex vivo with paroxetine or vehicle for 48 hours. In PBMCs, the concentration used for paroxetine was chosen to match therapeutic concentrations in the serum according to the consensus guidelines for therapeutic drug monitoring in psychiatry (51). We found that higher amounts of FKBP51 were significantly associated with smaller paroxetine-induced changes in phosphorylation of DNMT1 (Fig. 5D), further corroborating the notion that FKBP51 modulates the effects of paroxetine on DNMT1 phosphorylation.

Because FKBP51 abundance is increased upon stress (28–30), we explored the possibility that stress-induced glucocorticoids change pDNMT1 through increasing the amount of FKBP51. PBMCs were isolated from 21 healthy individuals before and after oral intake of dexamethasone, a syn-

thetic corticosteroid that selectively activates GR. In response to dexamethasone, we observed a negative correlation between the change in the phosphorylation of DNMT1 and the change in FKBP51 abundance: after dexamethasone treatment, increased abundance of FKBP51 was correlated with decreased phosphorylation of DNMT1 (Fig. 5E), consistent with the role of FKBP51 as mediator of stress-induced dephosphorylation of DNMT1.

Because higher protein amounts of FKBP51 in PBMCs were associated with better clinical treatment outcome in depressed patients (33), we hypothesized that phosphorylation of DNMT1 might also be linked to treatment response. Therefore, we determined the level of phosphorylation of DNMT1 in PBMCs from 40 patients of the Munich Antidepressant Response Signature (MARS) (52) project before and after 6 weeks of psychopharmacological treatment (with various antidepressants by doctor's choice). The change of phosphorylation of DNMT1 showed a negative correlation with the clinical treatment response [presented as the reduction in the Hamilton Depression Rating Scale (HDRS) from beginning of treatment to 6 weeks later] (Fig. 6A). Similarly, when we collected PBMCs from the patients at the time of admission to the clinic and treated the cells with paroxetine ex vivo, the cellular response (a decrease) in the phosphorylation of DNMT1 also negatively correlated with patients' (increased) reduction in HDRS (Fig. 6B). The change in the phosphorylation of DNMT1 observed in PBMCs in response to paroxetine ex vivo and the change in the phosphorylation of DNMT1 observed in patients after clinical treatment were well correlated (Fig. 6C). Overall, nonresponders tended to exhibit an increase in the phosphorylation of DNMT1 after 6 weeks of clinical treatment or after paroxetine treatment of their PBMCs ex vivo (fig. S8, A and B). When determining BDNF abundance, we found that the increase in BDNF secreted from PBMCs cultured and treated with paroxetine ex vivo significantly correlated with the abundance of FKBP51 (Fig. 6D). Likewise, the change in BDNF concentration in the serum of patients 6 weeks after clinical treatment also positively correlated with the abundance of FKBP51 in their blood cells collected at the beginning of treatment (Fig. 6E). Furthermore, the BDNF response in PBMCs cultured with paroxetine ex vivo positively correlated with the clinical treatment outcome observed in patients (Fig. 6F). A trend toward positive correlation was also observed for the change of BDNF in the serum and the clinical treatment outcome after 6 weeks (Fig. 6G). Together, these results support the physiological and clinical relevance of FKBP51 in inhibiting DNMT1 activation and promoting BDNF expression in modulating the action of antidepressants.

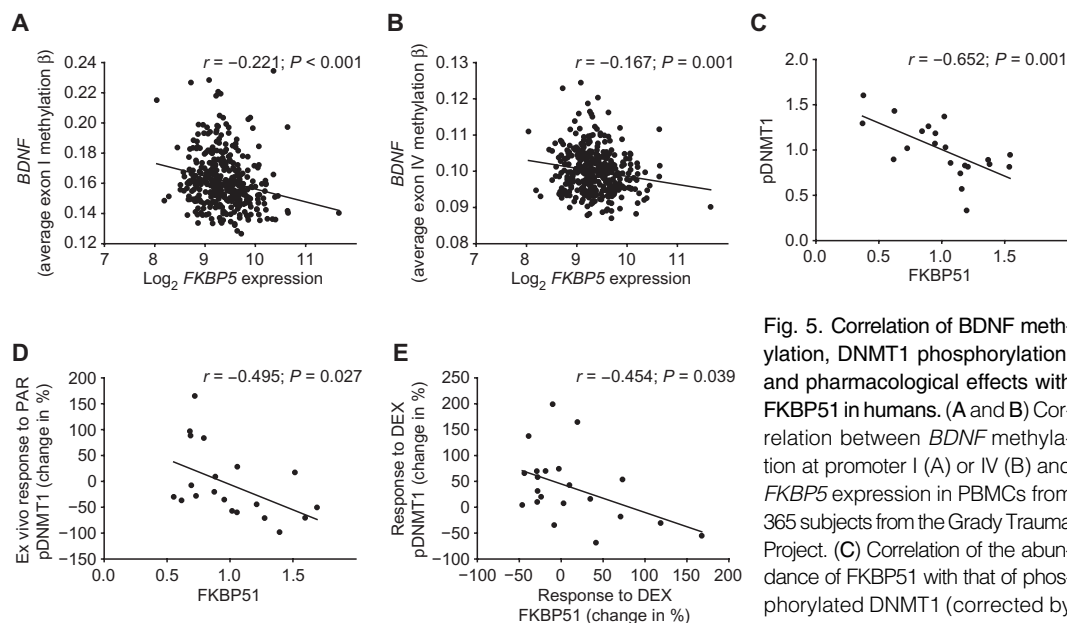


Fig. 5. Correlation of BDNF methylation, DNMT1 phosphorylation, and pharmacological effects with FKBP51 in humans. (A and B) Correlation between BDNF methylation at promoter I (A) or IV (B) and FKBP5 expression in PBMCs from 365 subjects from the Grady Trauma Project. (C) Correlation of the abundance of FKBP51 with that of phosphorylated DNMT1 (corrected by total DNMT1) in PBMCs from

healthy individuals ($n = 21$). (D) Correlation of FKBP51 abundance with the change in pDNMT1 abundance after ex vivo paroxetine exposure (48 hours, relative to vehicle) in PBMCs from healthy individuals ($n = 20$). (E) Correlation of the change in the phosphorylation of DNMT1 with that of FKBP51 in PBMCs isolated from healthy individuals that received dexamethasone (DEX, 1.5 mg, 6 hours; $n = 21$). Data were analyzed by Pearson correlation (A and B) or partial correlation corrected for age and gender (C to E).

DISCUSSION

Adaptation to stressful life events is a fundamental physiological process that involves several mechanisms, including epigenetic programming (24, 53, 54). Gene

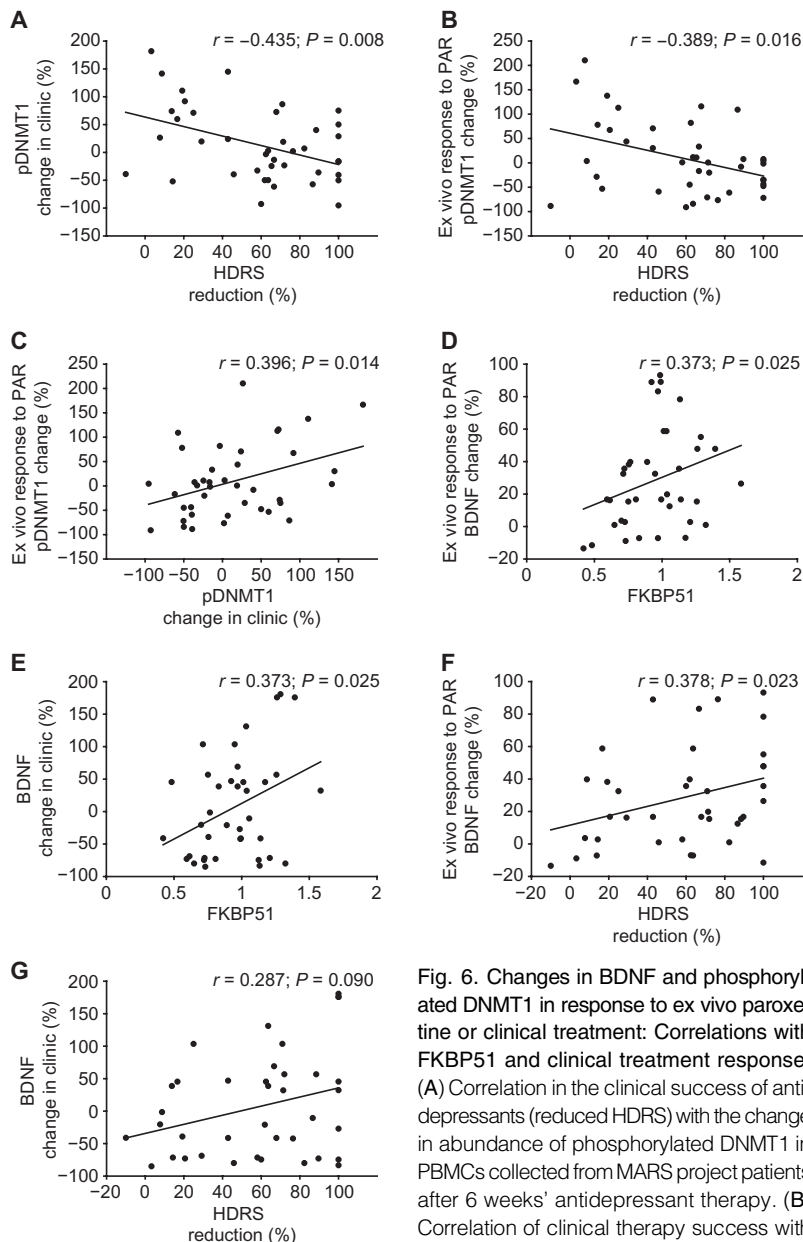


Fig. 6. Changes in BDNF and phosphorylated DNMT1 in response to ex vivo paroxetine or clinical treatment: Correlations with FKBP51 and clinical treatment response. (A) Correlation in the clinical success of antidepressants (reduced HDRS) with the change in abundance of phosphorylated DNMT1 in PBMCs collected from MARS project patients after 6 weeks' antidepressant therapy. (B) Correlation of clinical therapy success with the change in the phosphorylation of DNMT1

in respective patients' PBMCs isolated before therapy and treated ex vivo with paroxetine. (C) Correlation of the change in the phosphorylation of DNMT1 in PBMCs isolated from patients after clinical therapy [as described in (A)] with that in respective patients' PBMCs isolated before therapy and treated ex vivo with paroxetine. (D and E) Correlation of the abundance of FKBP51 with that of BDNF in PBMCs either (D) isolated before therapy and treated ex vivo with paroxetine or (E) isolated after clinical therapy [as described in (A)]. (F and G) Correlation of the clinical success of antidepressants (reduced HDRS) with that of BDNF in PBMCs either (F) isolated before therapy and treated ex vivo with paroxetine or (G) isolated after clinical therapy [as described in (A)]. Protein abundance was assessed in cell extracts by Western blotting. The abundance of phosphorylated DNMT1 was normalized to total DNMT1. Data were analyzed by Pearson correlation coefficient corrected for age and gender.

programming through the GR is evident from several reports (55). The established paradigm is that GR binds to chromatin targets and interacts with the epigenetic machinery at these sites, thereby changing local epigenetic

marks. Our study adds another twist to GR-mediated epigenetic programming by delineating a new route: we found that the GR-enhanced cochaperone FKBP51 suppresses the activity of the epigenetic enzyme DNMT1 by impairing its formation of a heterocomplex with CDK5. Our data suggest that FKBP51 displaces FKBP52 from CDK5, thereby preventing the subsequent interaction with and phosphorylation of DNMT1 at Ser¹⁵⁴, causing reduced activity of DNMT1 and decreased DNA methylation in the genome, including at the *Bdnf* gene. This displacement favoring FKBP51 chaperoning is promoted by paroxetine and possibly other antidepressants (Fig. 7).

Part of the translational aspect of this study uses PBMCs cultivated and treated ex vivo. Although signal transduction in peripheral cells ex vivo may not reliably replicate molecular activity in brain cells, the proteins involved in this study are present in multiple cell types and appear to have similar functional interactions in human PBMCs as they did in mouse brain tissue and astrocytes. Furthermore, our correlation analyses indicated that examining the effects on the activation of DNMT1 in patient PBMCs ex vivo may serve as a biomarker to predict the clinical response to antidepressants.

The data reported here support the notion that the role of chaperones in the management of stress reactivity extends beyond protein homeostasis to the genome and epigenome. For example, several landmark studies from the Lindquist laboratory document the role of Hsp90 in the evolution of heritable new traits in several organisms (56–60). The new traits shaped not only the phenotype but also the responsiveness to diverse drugs.

Our finding of the association of the chaperone FKBP51 with CDK5 in the regulation of DNMT1 adds insight into the molecular and physiological functions of FKBP51, which appears to be a versatile protein. In addition to its role as a potent inhibitor of GR (25–27), FKBP51 regulates other steroid receptors and associates with and regulates the kinase AKT through the recruitment of PH domain and leucine-rich repeat protein phosphatases (PHLPPs), with implications for cancer treatment (61). FKBP51 also forms protein complexes with Beclin1, AKT, and PHLPPs in the regulation of autophagy and affects tau stability, microtubule polymerization, neurite outgrowth, glycogen synthase kinase 3 β (GSK3 β) signaling, aging, and nuclear factor κ B (NF κ B) signaling in immune processes (26, 33, 36, 41, 62–64). Similar to the divergent effects of FKBP51 and FKBP52 on CDK5, the two highly homologous FKBP5s are reported to exert opposite effects on several other cellular processes (64). Even though both FKBP5s feature a well-described enzymatic activity [peptidylprolyl isomerization (65)], it appears that they rather function as protein scaffolds to promote various and at least partially differing protein complexes, such as what we observed regarding their interaction with CDK5.

Several functions of FKBP51 appear to contribute to its role in the cellular and organismal response to antidepressants (31–35). Originally, because of its effect on GR signaling (25), FKBP51 was included as candidate in the first gene association study in depression (31). Since then, cellular, animal, and clinical data suggest that the role of FKBP51 in regulating GSK3 β , AKT, and autophagy signaling mediate antidepressant effects (33, 36). GSK3 β and AKT decrease or increase, respectively, DNMT1 protein abundance (66). In our data set, we found no evidence that the presence

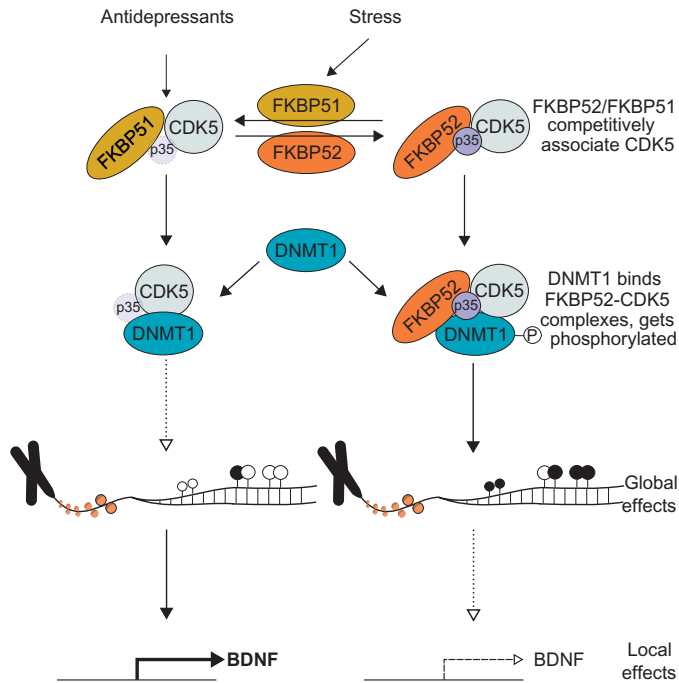


Fig. 7. Model of antidepressant effects on global and local epigenetic marks mediated through FKBP51-regulated suppression of DNMT1 activity. Schematic summarizing of our findings. FKBP51 competes with FKBP52 for binding to CDK5 and its activator p35. Because DNMT1 preferentially associates with FKBP52-CDK5-p35 complexes, DNMT1 displays higher phosphorylation and activity in the presence of high FKBP52 abundance but not when FKBP51 dominates the CDK5-p35 complexes. This causes differences in global DNA methylation and *BDNF* promoter methylation and expression. The stress-enhanced protein FKBP51 may thus mediate the impact of stress on epigenetic programming. Furthermore, this pathway is targeted by paroxetine and possibly other antidepressants to swap chaperone binding to CDK5, which may mediate the clinical response in patients with depression.

of FKBP51 alters the overall abundance of DNMT1, possibly because FKBP51 inhibits both AKT and GSK3 β (33, 36, 61). Although we suggest that the specific suppression of DNMT1 phosphorylation is mediated through the interaction of FKBP51 with CDK5, given the potential for multiple kinases being affected by FKBP51 (37), we cannot yet exclude the possibility that kinases other than CDK5 contribute to this mechanism.

The present and several additional studies portray FKBP51 as “reactivity protein” in the sense that it shapes the responsiveness to stress and drug treatment (33, 36, 61, 67, 68). Part of this conclusion extends from experiments with 51KO mice. Deletion of genes often goes along with compensatory mechanisms that are evoked by the organism throughout development. The effect of *FKBP5* deletion is not always detectable in the absence of a stimulus (33, 36, 67, 68); in our study, a difference in phosphorylation of DNMT1 between brains from wild-type and those from 51KO animals was only evident after treatment with paroxetine. Thus, it appears that possible compensatory changes in response to gene deletion do not compromise the effects evoked by challenges later in life.

It is tempting to speculate about the possibility to target (enhance) FKBP51 or downstream processes for the treatment of depression or other stress-related diseases. However, we would like to point out that due to the multifactorial actions of FKBP51, more studies are necessary to dissect which functions of FKBP51 are important and how it might be possible

to specifically affect a subset of these functions when targeting FKBP51 directly rather than by targeting downstream processes. A promising FKBP51-targeting compound has been presented recently (69). This compound is designed as inhibitor of the peptidylprolyl isomerase activity but because this activity appears dispensable for many functions of FKBP51, it is not clear yet which of the molecular actions of FKBP51 are influenced in which direction. Because the gene encoding FKBP51 has also been suggested as a risk factor for psychiatric disorders (31, 49, 70), timing of pharmacological FKBP51 targeting will likely be important. Moreover, if one limits the ability of FKBP51 to reduce GR function, more FKBP51 might be produced because of an ultrashort feedback loop (28, 30, 71). Alternatively, our study also supports the consideration of DNMT1 inhibitors for drug development in depression (14, 48).

MATERIALS AND METHODS

Cells

HEK293 cells [American Type Culture Collection (ATCC), CRL-1573], SKNMC cells (gift from C. Behl, University of Mainz, Germany), and MEFs (gift from M. Cox, University of Texas at El Paso) were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal calf serum (FCS) and 100 U/ml penicillin and streptomycin, respectively. Enriched astroglial cultures were prepared from postnatal day-1 Sprague-Dawley rat pups (Charles River) or wild-type and 51KO mice and were handled as described previously (72).

Transfection of astrocytes and MEF cells

Detached MEFs or cortical astrocytes (2×10^6) were resuspended in 100 μ l of transfection buffer [50 mM Hepes (pH 7.3), 90 mM NaCl, 5 mM KCl, and 0.15 mM CaCl $_2$] (73). Up to 5 μ g of plasmid DNA expressing the respective construct was added to the cell suspension, and electroporation was carried out using the Amaxa Nucleofactor system (Lonza). Cells were replated at a density of $10^5 \times \text{cm}^{-2}$ and further processed for Western blot analysis or assessment of DNMT1 activity.

Plasmids

The constructs expressing FLAG-tagged FKBP51, PPIase-deficient FKBP51 (F67D/D68V) or FKBP52, and hemagglutinin-tagged CDK5 have been described previously (25, 74). The plasmids pcDNA3/Myc-DNMT3A and pcDNA3/Myc-DNMT3B1 were from Addgene (#35521 and #35522).

Coimmunoprecipitation

Coimmunoprecipitations (CoIPs) of FLAG-tagged FKBP51/52 or endogenous CDK5, FKBP51, and FKBP52 were performed in HEK293 cells essentially as described previously (26). Briefly, 5×10^6 cells were electroporated with 5 μ g of the respective expression plasmids using a GenePulser (Bio-Rad) at 350 V/700 μ F in 400 μ l of electroporation buffer [50 mM K $_2$ HPO $_4$ /KH $_2$ PO $_4$, 20 mM KAc (pH 7.35), and 25 mM MgSO $_4$]. After 3 days of cultivation in DMEM/10% FCS, cells were lysed in CoIP buffer [20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA, and 0.5% Igepal complemented with protease inhibitor cocktail (Sigma)] for 20 min at 4°C with constant mixing. In the case of precipitating endogenous proteins from brain lysates, the tissue was homogenized and lysed in the same buffer. The lysates were cleared by centrifugation, and the protein concentration was determined and adjusted (brain lysates, 1 μ g $\times \mu$ l $^{-1}$; cell lysates, 1.2 μ g $\times \mu$ l $^{-1}$); 1 ml of lysate was incubated with 2.5 μ g of FLAG, FKBP51, FKBP52, or CDK5 antibody overnight at 4°C with constant mixing. Subsequently, 20 μ l of bovine serum albumin-blocked protein G Dynabeads (Invitrogen, 100-03D) were added to the lysate-antibody mix followed by

a 3-hour incubation at 4°C. Beads were washed three times with PBS, and bound proteins were eluted with 100 µl of 1 × FLAG peptide solution (100 to 200 µg × ml⁻¹, Sigma F3290) in PBS for 30 min at 4°C. In case of precipitation of endogenous proteins, elution was performed by adding 60 µl of Laemmli sample buffer and by incubation at 95 °C for 5 min. Five to fifteen micrograms of the input lysates or 2.5 to 5 µl of the immunoprecipitates were separated by SDS–polyacrylamide gel electrophoresis (PAGE) and analyzed by Western blotting. When quantifying coimmunoprecipitated proteins, their signals were normalized to input protein and to precipitated interactor protein.

Western blot analysis

Western blot analysis was conducted as previously described (13). Briefly, protein extracts were obtained by lysing cells in 62.5 mM Tris, 2% SDS, and 10% sucrose, supplemented with protease (Sigma, P2714) and phosphatase (Roche, 04906837001) inhibitor cocktail, followed by sonication of samples and heating at 95°C for 5 min. Proteins were separated by SDS-PAGE and electrotransferred onto nitrocellulose membranes. Blots were placed in Tris-buffered saline (TBS) supplemented with 0.05% Tween (Sigma, P2287) and 5% nonfat milk for 1 hour at room temperature, followed by an incubation with the primary antibody (diluted in TBS/0.05% Tween) overnight at 4°C. Primary antibodies recognizing the following epitopes or proteins were used: FLAG (1:7000; Rockland, 600-401-383), FKBP51 (1:1000; Bethyl, A301-430A), FKBP52 (1:2000; Bethyl, A301-427A), actin (1:5000; Santa Cruz Biotechnology, sc-1616), DNMT1 (1:1000; Imgenex, IMG-261A), CDK5 (1:1000; Cell Signaling Technology, #2506), p25 and p35 (1:1000; Cell Signaling Technology, #2680), p39 (1:1000; Cell Signaling Technology, #3275), myc (1:1000; Sigma-Aldrich, C3956), and Hsc70 (heat-shock cognate 70; 1:2000; Santa Cruz Biotechnology, sc-7298). Subsequently, blots were washed and probed with the respective horseradish peroxidase–conjugated secondary antibody for 1 hour at room temperature. Enhanced chemiluminescence detection reagent (Millipore) was applied to visualize the immunoreactive bands at ChemiDoc MP (Millipore). In the figures, bands corresponding to the respective proteins are displayed. Full lane blots corresponding to Figs. 1 to 3 and Figs. S1 and S3 are shown in data files S1 to S3.

In situ hybridization

Mouse tissue was available from experiments described previously (33). Frozen brains were coronally sectioned in a cryostat microtome at 18 µm and kept at –80°C. In situ hybridization using a ³⁵S uridine triphosphate–labeled ribonucleotide probe for BDNF (forward primer: 5′-GCGGCAGATAAAAAGACTGC and reverse primer: 5′-AAGTTGTGCGCAATGACTG; size, 495 bp) was performed as described previously (75). The slides were exposed to Kodak BioMax magnetic resonance films (Eastman Kodak Co.) and developed. Autoradiographs were digitized, and expression was determined by optical densitometry using the freely available National Institutes of Health (NIH) ImageJ software. The mean of two unilateral measurements (dorsal hippocampal subregions: CA1, CA2, CA3, and dentate gyrus) was calculated for each animal, subtracting the background signal of a nearby structure not expressing the gene of interest from the measurements.

Determination of BDNF

BDNF was quantified in the serum using the commercially available enzyme-linked immunosorbent assay (ELISA) kit human BDNF DuoSet, (R&D Systems, #DY248).

Subjects and preparation of human PBMCs

Human PBMCs were collected from 21 healthy male volunteers (average age was 25.8 ± 2.7 years) for analysis of protein–protein correlations

(Fig. 5C). The same volunteers received 1.5 mg of dexamethasone (orally), and protein expression changes in PBMCs were determined 6 hours later (Fig. 5E). PBMCs from another group of volunteers (20 men, average age of 34.8 ± 6.9 years) were collected for ex vivo cultivation and determination of paroxetine effects (Fig. 5D). In addition, PBMCs were collected before and after antidepressant treatment from 40 patients participating in the MARS study (52) and diagnosed with depression according to the diagnostic criteria of Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV; 23 women and 17 men, average age of 48.85 ± 14.7 years).

MARS is an open-label trial investigating outcome predictors for antidepressant treatment. The type of treatment was chosen according to the attending doctor's choice and in agreement with the patients. Dosage was adjusted and monitored according to plasma medication concentrations. Treatment outcome was weekly evaluated with the 21 items version of the HDRS. Fasting venous blood samples were collected through venipuncture on admission and after 6 weeks of antidepressant treatment. Samples were diluted with PBS, carefully loaded on Biocoll solution (Biochrom AG, L6113) and centrifuged at 800g for 20 min. PBMCs were enriched by selecting the interphase of the Biocoll gradient, followed by washing two times with ice-cold PBS. Cells were then resuspended in RPMI and plated at 4 × 10⁵/cm². After recovery for 6 hours, cells were treated with either 365 nM paroxetine or vehicle. This concentration has been chosen to match therapeutic concentrations in the serum according to the consensus guidelines for therapeutic drug monitoring in psychiatry (51). Patients and healthy subjects gave informed written consent, and the study was approved by the ethics committee of the Ludwig Maximilians University in Munich, Germany.

Global methylation analysis

Total cytosine methylation and hydroxymethylation in genomic DNA from rat astrocytes was determined by quantitative LC/UV-ESI-MS/MS analysis of digested DNA samples as described previously (76).

Gene expression and DNA methylation profiling in subjects from the Grady Trauma Project

Whole-blood samples were obtained from 365 subjects from the Grady Trauma Project. The Grady Trauma Project is an ongoing study that includes more than 6000 subjects from a highly traumatized, urban population of low socioeconomic status and examines the role of genetic and environmental factors on stress responses (49, 50). All subjects provided written informed consent, and all procedures were approved by the Institutional Review Boards of the Emory University School of Medicine and Grady Memorial Hospital.

To assess DNA methylation, whole-blood genomic DNA was extracted using the Gentra Puregene Blood Kit (Qiagen). DNA quantity and quality were assessed by Quant-iT PicoGreen (Invitrogen) and NanoDrop 2000 Spectrophotometer (Thermo Scientific). Subsequently, DNA was bisulfite-converted with the Zymo EZ-96 DNA Methylation Kit (Zymo Research), and DNA methylation was assessed with Illumina HumanMethylation450 BeadChip (450K) arrays. Hybridization and processing was performed as previously described (77). Quality control of methylation data, including intensity readouts, normalization, and cellular composition estimation, was carried out using the minfi Bioconductor R package version 1.10.2 (78). Failed probes were excluded on the basis of a detection of *P* value greater than 0.01 in at least 75% of the samples (*n* = 233 probes). We also removed probes in X or Y chromosome and nonspecific binding probes (79) if single nucleotide polymorphisms (SNPs) were documented in the interval for which the Illumina probe is designed to hybridize or if they were located close (10 bp from query site) to SNPs reported in the 1000 Genomes Project to have minor allele frequency ≥ 0.05 (80). Data were normalized

with functional normalization included in the minfi R package (81). Batch effects were identified after inspection of principal component analysis using the shinyMethyl Bioconductor R package version 0.99.39 and removed using COMBAT (82). As previously recommended (83), we used *M* values to perform the statistical analyses involving DNA methylation and β methylation values to visualize the relationships with gene expression.

To assess gene expression, whole-blood RNA was collected with PAX-gene Blood RNA Tubes (PreAnalytiX) and processed as previously described (84). Blood RNA was then hybridized to Illumina HumanHT-12 version 3 and version 4 Expression BeadChips (Illumina). Gene expression was measured using the Illumina HumanHT-12 version 3 Expression BeadChip Kit (Illumina). Raw microarray scan files were exported with the Illumina BeadStudio program 13 and were analyzed with R (www.R-project.org). Using Illumina internal controls, microarray data were transformed and normalized through variance stabilizing normalization (85). Potential confounding as a result of batch effects was corrected using an empirical Bayes method (82). Data reproducibility was confirmed with six pairs of technical replicates (average Pearson correlation = 0.996). The raw array data for the Grady study have been deposited to Gene Expression Omnibus (GEO) both for gene expression (GSE58137) and DNA methylation (GSE72680).

DNMT1 activity assay

DNMT1 activity was measured as previously described (13), with the modification that astrocytes were also transfected with FKBP51, FKBP52, or vector plasmids 3 days before cell harvesting.

Bdnf promoter methylation analysis

The methylation status of *Bdnf* promoters I and IV in mouse brain was analyzed by bisulfite pyrosequencing. Briefly, total DNA was isolated from specific brain regions using NucleoSpin Tissue (Macherey-Nagel), according to the manufacturer's instructions. After quantification, about 300 ng of DNA were bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research), and bisulfite-converted DNA samples were used as templates for polymerase chain reactions (PCRs) amplifying promoters I and IV of the *Bdnf* gene. Primers used for the bisulfite PCR and pyrosequencing are provided in table S4. Pyrosequencing primers were designed with the MethMarker software and carried out on a PSQ96 (Pyrosequencing) using PyroMark Gold Q96 reagents (Qiagen) according to the manufacturer's recommendations.

Statistical analysis

Statistical analyses were performed with SigmaPlot 13.0 and SPSS 18. Student's *t* tests or Mann-Whitney tests were applied to compare two groups, whereas one- or two-way analysis of variance (ANOVA) were performed for comparisons between three or more groups, followed by Bonferroni, Tukey's, or Duncan's post hoc test, as appropriate. Correlations between variables were analyzed using the Pearson correlation coefficient. Treatment outcome in the patient sample was determined as percent reduction of the HDRS rating scores between admission and after 6 weeks of antidepressant treatment. In case of an early discharge from the hospital (after at least 2 weeks of antidepressant treatment), missing HDRS scores were estimated using nonlinear regression with sex, age, and previous HDRS scores as predictor variables. *P* values lower than 0.05 were considered statistically significant.

Ethics statement

All experiments were carried out in the animal facilities of the Max Planck Institute of Psychiatry in Munich, Germany. The experiments were carried out in accordance with the European Communities' Council Directive 86/609/EEC. All efforts were made to minimize animal suffering during

the experiments. The protocols were approved by the committee for the Care and Use of Laboratory Animals of the Government of Upper Bavaria, Germany. Approval for the MARS project was received by the ethics committee in charge (submission no. 318/00, ethics committee of the Medical Faculty at the Ludwig Maximilians University, Munich, Germany), and participants gave oral and written consent before study inclusion.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/8/404/ra119/DC1

Fig. S1. DNMT protein interaction analysis.

Fig. S2. FKBP51 and FKBP52 differentially affect the phosphorylation and activity of DNMT1.

Fig. S3. Paroxetine promotes FKBP51 binding to CDK5.

Fig. S4. FKBP51 shapes the effect of paroxetine on DNMT1.

Fig. S5. The effect of amitriptyline and haloperidol on the phosphorylation of DNMT1.

Fig. S6. The effect of amitriptyline on *Bdnf* promoter IV methylation.

Fig. S7. DNA methylation of *Bdnf* inversely correlates with the abundance of *FKBP5* mRNA.

Fig. S8. Cellular and clinical change in the abundance of phosphorylated DNMT1 in PBMCs in response to antidepressants.

Table S1. Pyrosequencing results of *Bdnf* promoter methylation analyses in wild-type and 51KO mice.

Table S2. Statistical details of ANOVA analysis by figure panel.

Table S3. Statistical details of Student's *t* test or Mann-Whitney analysis by figure panel.

Table S4. Primers for bisulfite PCR and pyrosequencing.

Data file S1. Blot collections for Fig. 1 and figs. S1 and S3.

Data file S2. Blot collections for Fig. 2.

Data file S3. Blot collections for Fig. 3.

REFERENCES AND NOTES

1. T. Klengel, E. B. Binder, Gene-environment interactions in major depressive disorder. *Can. J. Psychiatry* **58**, 76–83 (2013).
2. F. J. Raabe, D. Spengler, Epigenetic risk factors in PTSD and depression. *Front. Psychiatry* **4**, 80 (2013).
3. E. J. Nestler, Epigenetic mechanisms of depression. *JAMA Psychiatry* **71**, 454–456 (2014).
4. S. L. Berger, The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412 (2007).
5. R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, Structure and function of mammalian DNA methyltransferases. *Chembiochem* **12**, 206–222 (2011).
6. T. Klengel, E. B. Binder, Gene \times environment interactions in the prediction of response to antidepressant treatment. *Int. J. Neuropsychopharmacol.* **16**, 701–711 (2013).
7. V. Vialou, J. Feng, A. J. Robison, E. J. Nestler, Epigenetic mechanisms of depression and antidepressant action. *Annu. Rev. Pharmacol. Toxicol.* **53**, 59–87 (2013).
8. A. Menke, E. B. Binder, Epigenetic alterations in depression and antidepressant treatment. *Dialogues Clin. Neurosci.* **16**, 395–404 (2014).
9. N. M. Tsankova, O. Berton, W. Renthal, A. Kumar, R. L. Neve, E. J. Nestler, Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nat. Neurosci.* **9**, 519–525 (2006).
10. T. Perisic, N. Zimmermann, T. Kirmeier, M. Asmus, F. Tuorto, M. Uhr, F. Holsboer, T. Rein, J. Zschocke, Valproate and amitriptyline exert common and divergent influences on global and gene promoter-specific chromatin modifications in rat primary astrocytes. *Neuropsychopharmacology* **35**, 792–805 (2010).
11. M. Fatemi, A. Hermann, H. Gowher, A. Jeltsch, Dnmt3a and Dnmt1 functionally co-operate during de novo methylation of DNA. *Eur. J. Biochem.* **269**, 4981–4984 (2002).
12. M. G. Goll, T. H. Bestor, Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
13. N. Zimmermann, J. Zschocke, T. Perisic, S. Yu, F. Holsboer, T. Rein, Antidepressants inhibit DNA methyltransferase 1 through reducing G9a levels. *Biochem. J.* **448**, 93–102 (2012).
14. A. J. Sales, C. Biojone, M. S. Terceti, F. S. Guimarães, M. V. M. Gomes, S. R. L. Joca, Antidepressant-like effect induced by systemic and intra-hippocampal administration of DNA methylation inhibitors. *Br. J. Pharmacol.* **164**, 1711–1721 (2011).
15. A. Tadić, L. Müller-Engling, K. F. Schlicht, A. Kotsiari, N. Dreimüller, A. Kleimann, S. Bleich, K. Lieb, H. Frieling, Methylation of the promoter of brain-derived neurotrophic factor exon IV and antidepressant response in major depression. *Mol. Psychiatry* **19**, 281–283 (2014).
16. Y. Dwivedi, H. S. Rizavi, G. N. Pandey, Antidepressants reverse corticosterone-mediated decrease in brain-derived neurotrophic factor expression: Differential regulation of specific exons by antidepressants and corticosterone. *Neuroscience* **139**, 1017–1029 (2006).

17. B. G. Dias, S. B. Banerjee, R. S. Duman, V. A. Vaidya, Differential regulation of Brain Derived Neurotrophic Factor transcripts by antidepressant treatments in the adult rat brain. *Neuropharmacology* **45**, 553–563 (2003).
18. W. Qin, H. Leonhardt, G. Pichler, Regulation of DNA methyltransferase 1 by interactions and modifications. *Nucleus* **2**, 392–402 (2011).
19. S. Kar, M. Deb, D. Sengupta, A. Shilpi, S. Parbin, J. Torrisani, S. Pradhan, S. Patra, An insight into the various regulatory mechanisms modulating human DNA methyltransferase 1 stability and function. *Epigenetics* **7**, 994–1007 (2012).
20. J. F. Glickman, J. G. Pavlovich, N. O. Reich, Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. *J. Biol. Chem.* **272**, 17851–17857 (1997).
21. R. Goyal, P. Rathert, H. Laser, H. Gowher, A. Jeltsch, Phosphorylation of serine-515 activates the Mammalian maintenance methyltransferase Dnmt1. *Epigenetics* **2**, 155–160 (2007).
22. G. Lavoie, P.-O. Estève, N. B. Laulan, S. Pradhan, Y. St-Pierre, PKC isoforms interact with and phosphorylate DNMT1. *BMC Biol.* **9**, 31 (2011).
23. G. Lavoie, Y. St-Pierre, Phosphorylation of human DNMT1: Implication of cyclin-dependent kinases. *Biochem. Biophys. Res. Commun.* **409**, 187–192 (2011).
24. E. R. de Kloet, M. Joëls, F. Holsboer, Stress and the brain: From adaptation to disease. *Nat. Rev. Neurosci.* **6**, 463–475 (2005).
25. G. M. Wochnik, J. Rüegg, G. A. Abel, U. Schmidt, F. Holsboer, T. Rein, FK506-binding proteins 51 and 52 differentially regulate dynein interaction and nuclear translocation of the glucocorticoid receptor in mammalian cells. *J. Biol. Chem.* **280**, 4609–4616 (2005).
26. J.-P. Schülke, G. M. Wochnik, I. Lang-Rollin, N. C. Gassen, R. T. Knapp, B. Berning, A. Yassouridis, T. Rein, Differential impact of tetratricopeptide repeat proteins on the steroid hormone receptors. *PLOS One* **5**, e11717 (2010).
27. D. L. Riggs, P. J. Roberts, S. C. Chirillo, J. Cheung-Flynn, V. Prapapanich, T. Ratajczak, R. Gaber, D. Picard, D. F. Smith, The Hsp90-binding peptidylprolyl isomerase FKBP52 potentiates glucocorticoid signaling in vivo. *EMBO J.* **22**, 1158–1167 (2003).
28. T. R. Hubler, J. G. Scammell, Intronic hormone response elements mediate regulation of FKBP5 by progestins and glucocorticoids. *Cell Stress Chaperones* **9**, 243–252 (2004).
29. R. S. Lee, K. L. K. Tamashiro, X. Yang, R. H. Purcell, A. Harvey, V. L. Willour, Y. Huo, M. Rongione, G. S. Wand, J. B. Potash, Chronic corticosterone exposure increases expression and decreases deoxyribonucleic acid methylation of *Fkbp5* in mice. *Endocrinology* **151**, 4332–4343 (2010).
30. V. Paakinaho, H. Makkonen, T. Jääskeläinen, J. J. Palvimäki, Glucocorticoid receptor activates poised FKBP51 locus through long-distance interactions. *Mol. Endocrinol.* **24**, 511–525 (2010).
31. E. B. Binder, D. Salyakina, P. Lichtner, G. M. Wochnik, M. Ising, B. Pütz, S. Papiol, S. Seaman, S. Lucae, M. A. Kohli, T. Nickel, H. E. Künzel, B. Fuchs, M. Majer, A. Pfennig, N. Kern, J. Brunner, S. Modell, T. Baghai, T. Deiml, P. Zill, B. Bondy, R. Rupprecht, T. Messer, O. Köhnlein, H. Dabitz, T. Brückl, N. Müller, H. Pfister, R. Lieb, J. C. Mueller, E. Löhmussaar, T. M. Strom, T. Bettecken, T. Meitinger, M. Uhr, T. Rein, F. Holsboer, B. Müller-Mysok, Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment. *Nat. Genet.* **36**, 1319–1325 (2004).
32. M. Lekman, G. Laje, D. Charney, A. J. Rush, A. F. Wilson, A. J. Sorant, R. Lipsky, S. R. Wisniewski, H. Manji, F. J. McMahon, S. Paddock, The FKBP5 gene in depression and treatment response—an association study in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Cohort. *Biol. Psychiatry* **63**, 1103–1110 (2008).
33. N. C. Gassen, J. Hartmann, J. Zschocke, J. Stephan, K. Hafner, A. Zellner, T. Kirmeyer, L. Kollmannsberger, K. V. Wagner, N. Dedic, G. Balsevich, J. M. Deussing, S. Kloiber, S. Lucae, F. Holsboer, M. Eder, M. Uhr, M. Ising, M. V. Schmidt, T. Rein, Association of FKBP51 with priming of autophagy pathways and mediation of antidepressant treatment response: Evidence in cells, mice, and humans. *PLOS Med.* **11**, e1001755 (2014).
34. G. Laje, R. H. Perlis, A. J. Rush, F. J. McMahon, Pharmacogenetics studies in STAR*D: Strengths, limitations, and results. *Psychiatr. Serv.* **60**, 1446–1457 (2009).
35. Y.-F. Zou, F. Wang, X.-L. Feng, W.-F. Li, J.-H. Tao, F.-M. Pan, F. Huang, H. Su, Meta-analysis of FKBP5 gene polymorphisms association with treatment response in patients with mood disorders. *Neurosci. Lett.* **484**, 56–61 (2010).
36. N. C. Gassen, J. Hartmann, A. S. Zannas, A. Kretschmar, J. Zschocke, G. Maccarrone, K. Hafner, A. Zellner, L. Kollmannsberger, K. V. Wagner, D. Mehta, S. Kloiber, C. W. Turck, S. Lucae, G. P. Chrousos, F. Holsboer, E. B. Binder, M. Ising, M. V. Schmidt, T. Rein, FKBP51 inhibits GSK3 β and augments the effects of distinct psychotropic medications. *Mol. Psychiatry* **10**, 1038/10.1038/mp.2015.38 (2015).
37. M. Taipale, G. Tucker, J. Peng, I. Krykbaeva, Z. Y. Lin, B. Larsen, H. Choi, B. Berger, A. C. Gingras, S. Lindquist, A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* **158**, 434–448 (2014).
38. Q. Zhou, A. T. Agoston, P. Atadja, W. G. Nelson, N. E. Davidson, Inhibition of histone deacetylases promotes ubiquitin-dependent proteasomal degradation of DNA methyltransferase 1 in human breast cancer cells. *Mol. Cancer Res.* **6**, 873–883 (2008).
39. W. Luo, F. Dou, A. Rodina, S. Chip, J. Kim, Q. Zhao, K. Moullick, J. Aguirre, N. Wu, P. Greengard, G. Chiosis, Roles of heat-shock protein 90 in maintaining and facilitating the neurodegenerative phenotype in tauopathies. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9511–9516 (2007).
40. J. P. Lopes, P. Agostinho, Cdk5: Multitasking between physiological and pathological conditions. *Prog. Neurobiol.* **94**, 49–63 (2011).
41. U. K. Jinwal, J. Koren III, S. I. Borysov, A. B. Schmid, J. F. Abisambra, L. J. Blair, A. G. Johnson, J. R. Johnson, C. L. Shults, J. C. O'Leary III, Y. Jin, J. Buchner, M. B. Cox, C. A. Dickey, The Hsp90 cochaperone, FKBP51, increases Tau stability and polymerizes microtubules. *J. Neurosci.* **30**, 591–599 (2010).
42. R. L. Barent, S. C. Nair, D. C. Carr, Y. Ruan, R. A. Rimerman, J. Fulton, Y. Zhang, D. F. Smith, Analysis of FKBP51/FKBP52 chimeras and mutants for Hsp90 binding and association with progesterone receptor complexes. *Mol. Endocrinol.* **12**, 342–354 (1998).
43. T. Aid, A. Kazantseva, M. Piirsoo, K. Palm, T. Timmusk, Mouse and rat *BDNF* gene structure and expression revisited. *J. Neurosci. Res.* **85**, 525–535 (2007).
44. F. Boule, D. L. van den Hove, S. B. Jakob, B. P. Rutten, M. Hamon, J. van Os, K.-P. Lesch, L. Lanfume, H. W. Steinbusch, G. Kenis, Epigenetic regulation of the BDNF gene: Implications for psychiatric disorders. *Mol. Psychiatry* **17**, 584–596 (2012).
45. K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan, Y. E. Sun, DNA methylation-related chromatin remodeling in activity-dependent *BDNF* gene regulation. *Science* **302**, 890–893 (2003).
46. F. D. Lubin, T. L. Roth, J. D. Sweatt, Epigenetic regulation of *bdnf* gene transcription in the consolidation of fear memory. *J. Neurosci.* **28**, 10576–10586 (2008).
47. A. E. West, W. G. Chen, M. B. Dalva, R. E. Dolmetsch, J. M. Kornhauser, A. J. Shaywitz, M. A. Takasu, X. Tao, M. E. Greenberg, Calcium regulation of neuronal gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11024–11031 (2001).
48. Y. Wei, P. A. Melas, G. Wegener, A. A. Mathé, C. Lavebratt, Antidepressant-like effect of sodium butyrate is associated with an increase in TET1 and in 5-hydroxymethylation levels in the *Bdnf* gene. *Int. J. Neuropsychopharmacol.* **18**, pyu032 (2014).
49. E. B. Binder, R. G. Bradley, W. Liu, M. P. Epstein, T. C. Deveau, K. B. Mercer, Y. Tang, C. F. Gillespie, C. M. Heim, C. B. Nemeroff, A. C. Schwartz, J. F. Cubells, K. J. Ressler, Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. *JAMA* **299**, 1291–1305 (2008).
50. C. F. Gillespie, B. Bradley, K. Mercer, A. K. Smith, K. Conneely, M. Gape, T. Weiss, A. C. Schwartz, J. F. Cubells, K. J. Ressler, Trauma exposure and stress-related disorders in inner city primary care patients. *Gen. Hosp. Psychiatry* **31**, 505–514 (2009).
51. C. Hiemke, P. Baumann, N. Bergemann, A. Conca, O. Dietmaier, K. Egberts, M. Fric, M. Gerlach, C. Greiner, G. Gründer, E. Haen, U. Havemann-Reinecke, S. E. Jaquenoud, H. Kirchherr, G. Laux, U. C. Lutz, T. Messer, M. J. Müller, B. Pfuhlmann, B. Rambeck, P. Riederer, B. Schoppek, J. Stingl, M. Uhr, S. Ulrich, R. Waschler, G. Zernig, AGNP Consensus Guidelines for Therapeutic Drug Monitoring in Psychiatry: Update 2011. *Pharmacopsychiatry* **44**, 195–235 (2011).
52. J. M. Hennings, T. Owashi, E. B. Binder, S. Horstmann, A. Menke, S. Kloiber, T. Dose, B. Wollweber, D. Spieler, T. Messer, R. Lutz, H. Künzel, T. Bierner, T. Pollmächer, H. Pfister, T. Nickel, A. Sonntag, M. Uhr, M. Ising, F. Holsboer, S. Lucae, Clinical characteristics and treatment outcome in a representative sample of depressed inpatients—Findings from the Munich Antidepressant Response Signature (MARS) project. *J. Psychiatr. Res.* **43**, 215–229 (2009).
53. A. Harris, J. Seckl, Glucocorticoids, prenatal stress and the programming of disease. *Horm. Behav.* **59**, 279–289 (2011).
54. N. Provençal, E. B. Binder, The effects of early life stress on the epigenome: From the womb to adulthood and even before. *Exp. Neurol.* **268**, 10–20 (2015).
55. V. G. Moisiadis, S. G. Matthews, Glucocorticoids and fetal programming part 2: Mechanisms. *Nat. Rev. Endocrinol.* **10**, 403–411 (2014).
56. L. E. Cowen, S. Lindquist, Hsp90 potentiates the rapid evolution of new traits: Drug resistance in diverse fungi. *Science* **309**, 2185–2189 (2005).
57. N. Rohner, D. F. Jarosz, J. E. Kowalko, M. Yoshizawa, W. R. Jeffery, R. L. Borowsky, S. Lindquist, C. J. Tabin, Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. *Science* **342**, 1372–1375 (2013).
58. L. Whitesell, S. Santagata, M. L. Mendillo, N. U. Lin, D. A. Proia, S. Lindquist, HSP90 empowers evolution of resistance to hormonal therapy in human breast cancer models. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18297–18302 (2014).
59. D. F. Jarosz, S. Lindquist, Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science* **330**, 1820–1824 (2010).
60. S. L. Rutherford, S. Lindquist, Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
61. H. Pei, L. Li, B. L. Fridley, G. D. Jenkins, K. R. Kalari, W. Lingle, G. Petersen, Z. Lou, L. Wang, FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. *Cancer Cell* **16**, 259–266 (2009).
62. J. C. O'Leary III, S. Dharma, L. J. Blair, S. Brady, A. G. Johnson, M. Peters, J. Cheung-Flynn, M. B. Cox, G. de Erausquin, E. J. Weeber, U. K. Jinwal, C. A. Dickey, A new anti-depressive strategy for the elderly: Ablation of FKBP5/FKBP51. *PLOS One* **6**, e24840 (2011).
63. H. R. Quintá, D. Maschi, C. Gomez-Sanchez, G. Piwien-Pilipuk, M. D. Galgiani, Subcellular rearrangement of hsp90-binding immunophilins accompanies neuronal differentiation and neurite outgrowth. *J. Neurochem.* **115**, 716–734 (2010).

64. C. L. Storer, C. A. Dickey, M. D. Galigiana, T. Rein, M. B. Cox, FKBP51 and FKBP52 in signaling and disease. *Trends Endocrinol. Metab.* **22**, 481–490 (2011).
65. F. Pirk, J. Buchner, Functional analysis of the hsp90-associated human peptidyl prolyl *cis/trans* isomerases FKBP51, FKBP52 and cyp40. *J. Mol. Biol.* **308**, 795–806 (2001).
66. L. Sun, H. Zhao, Z. Xu, Q. Liu, Y. Liang, L. Wang, X. Cai, L. Zhang, L. Hu, G. Wang, X. Zha, Phosphatidylinositol 3-kinase/protein kinase B pathway stabilizes DNA methyltransferase I protein and maintains DNA methylation. *Cell. Signal.* **19**, 2255–2263 (2007).
67. J. Hartmann, K. V. Wagner, C. Liebl, S. H. Scharf, X.-D. Wang, M. Wolf, F. Hausch, T. Rein, U. Schmidt, C. Touma, J. Cheung-Flynn, M. B. Cox, D. F. Smith, F. Holsboer, M. B. Müller, M. V. Schmidt, The involvement of FK506-binding protein 51 (FKBP5) in the behavioral and neuroendocrine effects of chronic social defeat stress. *Neuropharmacology* **62**, 332–339 (2012).
68. C. Touma, N. C. Gassen, L. Herrmann, J. Cheung-Flynn, D. R. Büll, I. A. Ionescu, J.-M. Heinzmann, A. Knapman, A. Siebertz, A.-M. Depping, J. Hartmann, F. Hausch, M. V. Schmidt, F. Holsboer, M. Ising, M. B. Cox, U. Schmidt, T. Rein, FK506 binding protein 5 shapes stress responsiveness: Modulation of neuroendocrine reactivity and coping behavior. *Biol. Psychiatry* **70**, 928–936 (2011).
69. S. Gaali, A. Kirschner, S. Cuboni, J. Hartmann, C. Kozany, G. Balsevich, C. Namendorf, P. Fernandez-Vizarra, C. Sippel, A. S. Zannas, R. Draenert, E. B. Binder, O. F. Almeida, G. Rührer, M. Uhr, M. V. Schmidt, C. Touma, A. Bracher, F. Hausch, Selective inhibitors of the FK506-binding protein 51 by induced fit. *Nat. Chem. Biol.* **11**, 33–37 (2014).
70. A. Roy, E. Gorodetsky, Q. Yuan, D. Goldman, M.-A. Enoch, Interaction of *FKBP5*, a stress-related gene, with childhood trauma increases the risk for attempting suicide. *Neuropsychopharmacology* **35**, 1674–1683 (2010).
71. T. Klengel, D. Mehta, C. Anacker, M. Rex-Haffner, J. C. Pruessner, C. M. Pariante, T. W. Pace, K. B. Mercer, H. S. Mayberg, B. Bradley, C. B. Nemeroff, F. Holsboer, C. M. Heim, K. J. Ressler, T. Rein, E. B. Binder, Allele-specific *FKBP5* DNA demethylation mediates gene–childhood trauma interactions. *Nat. Neurosci.* **16**, 33–41 (2013).
72. B. Franke, M. Figiel, J. Engele, CNS glia are targets for GDNF and neurturin. *Histochem. Cell Biol.* **110**, 595–601 (1998).
73. B. G. Schumann, P. Jutzi, I. Roditi, Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Mol. Biochem. Parasitol.* **175**, 91–94 (2011).
74. S. van den Heuvel, E. Harlow, Distinct roles for cyclin-dependent kinases in cell cycle control. *Science* **262**, 2050–2054 (1993).
75. M. V. Schmidt, V. Sterlemann, K. Ganea, C. Liebl, S. Alam, D. Harbich, M. Greetfeld, M. Uhr, F. Holsboer, M. B. Müller, Persistent neuroendocrine and behavioral effects of a novel, etiologically relevant mouse paradigm for chronic social stress during adolescence. *Psychoneuroendocrinology* **32**, 417–429 (2007).
76. T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotjarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
77. D. Mehta, T. Klengel, K. N. Conneely, A. K. Smith, A. Altmann, T. W. Pace, M. Rex-Haffner, A. Loeschner, M. Gonik, K. B. Mercer, B. Bradley, B. Müller-Myhsok, K. J. Ressler, E. B. Binder, Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8302–8307 (2013).
78. M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, R. A. Irizarry, Minfi: A flexible and comprehensive Bioconductor package for the analysis of Illumina DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
79. E. M. Price, A. M. Cotton, L. L. Lam, P. Farré, E. Emberly, C. J. Brown, W. P. Robinson, M. S. Kobor, Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**, 4 (2013).
80. Y.-A. Chen, M. Lemire, S. Choufani, D. T. Butcher, D. Grafodatskaya, B. W. Zanke, S. Gallinger, T. J. Hudson, R. Weksberg, Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209 (2013).
81. J.-P. Fortin, A. Labbe, M. Lemire, B. W. Zanke, T. J. Hudson, E. J. Fertig, C. M. Greenwood, K. D. Hansen, Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
82. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
83. W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**, S96–S104 (2002).
84. P. Du, X. Zhang, C.-C. Huang, N. Jafari, W. A. Kibbe, L. Hou, S. M. Lin, Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
85. A. Menke, J. Arloth, B. Pütz, P. Weber, T. Klengel, D. Mehta, M. Gonik, M. Rex-Haffner, J. Rubel, M. Uhr, S. Lucae, J. M. Deussing, B. Müller-Myhsok, F. Holsboer, E. B. Binder, Dexamethasone stimulated gene expression in peripheral blood is a sensitive marker for glucocorticoid receptor resistance in depressed patients. *Neuropsychopharmacology* **37**, 1455–1464 (2012).

Acknowledgments: We are indebted to J. Hornung for performing the BDNF ELISA analysis. We also thank C. Behl (University of Mainz, Germany) and M. Cox (University of Texas at El Paso) for providing us with SKNMC cells and MEFs, respectively. **Funding:** N.C.G. was supported in part by a European Research Council (ERC) starting grant to E.B.B. [Framework Programme 7 grant no. 281338 (GxE-MOLMECH)]. A.S.Z. was supported in part by ERC starting grant 281338 and in part by a Marie-Sklodowska Curie fellowship (H2020 grant #653240). J.R. credits funding from the Swiss National Science Foundation (PZ00P3_126319) and from the Swedish Research Council Formas. G.P.C. was supported by the National and Kapodistrian University of Athens. **Author contributions:** N.C.G. and G.R.F. performed biochemical and molecular biology experiments and contributed to the study design and manuscript preparation; A.S.Z. performed the correlation analysis in the Grady sample and performed the statistical analyses; J.H. provided animal tissue and performed the in situ hybridization assays; J.Z. contributed to the study design; K.H. performed molecular biology and cell culture experiments; T.C.-R. and M.I. performed the statistical analyses; J.S. performed the global methylation analysis; S.N.P. contributed to the DNMT1 activity assay; L.H. contributed to the in situ hybridization assays; M.K. and F.W. provided the BDNF ELISA kit; S.K. and S.L. contributed to the organization of the MARS sample; G.P.C. contributed to manuscript preparation; T.C. performed the global DNA methylation assays; E.B.B. contributed to organizing the Grady sample collection; M.V.S. provided animal tissue; J.R. performed the bisulfite sequencing, primer design, and pyrosequencing; T.R. contributed to the study design and manuscript preparation. **Competing interests:** The authors declare they have no competing interests. **Data and materials availability:** The raw array data for the Grady study have been deposited to Gene Expression Omnibus, both for gene expression (GSE58137) and DNA methylation (GSE72680).

Submitted 10 June 2015

Accepted 6 November 2015

Final Publication 24 November 2015

10.1126/scisignal.aac7695

Citation: N. C. Gassen, G. R. Fries, A. S. Zannas, J. Hartmann, J. Zschocke, K. Hafner, T. Carrillo-Roa, J. Steinbacher, S. N. Preißinger, L. Hoeijmakers, M. Knop, F. Weber, S. Kloiber, S. Lucae, G. P. Chrousos, T. Carell, M. Ising, E. B. Binder, M. V. Schmidt, J. Rüegg, T. Rein, Chaperoning epigenetics: FKBP51 decreases the activity of DNMT1 and mediates epigenetic effects of the antidepressant paroxetine. *Sci. Signal.* **8**, ra119 (2015).

Chaperoning epigenetics: FKBP51 decreases the activity of DNMT1 and mediates epigenetic effects of the antidepressant paroxetine

Nils C. Gassen, Gabriel R. Fries, Anthony S. Zannas, Jakob Hartmann, Jürgen Zschocke, Kathrin Hafner, Tania Carrillo-Roa, Jessica Steinbacher, S. Nicole Preißinger, Lianne Hoeijmakers, Matthias Knop, Frank Weber, Stefan Kloiber, Susanne Lucae, George P. Chrousos, Thomas Carell, Marcus Ising, Elisabeth B. Binder, Mathias V. Schmidt, Joëlle Rüegg and Theo Rein (November 24, 2015)
Science Signaling **8** (404), ra119. [doi: 10.1126/scisignal.aac7695]

The following resources related to this article are available online at <http://stke.sciencemag.org>.
This information is current as of August 15, 2016.

Article Tools Visit the online version of this article to access the personalization and article tools:
<http://stke.sciencemag.org/content/8/404/ra119>

Supplemental Materials "Supplementary Materials"
<http://stke.sciencemag.org/content/suppl/2015/11/20/8.404.ra119.DC1>

Related Content The editors suggest related resources on *Science's* sites:
<http://stke.sciencemag.org/content/sigtrans/8/404/pc29.full>
<http://stke.sciencemag.org/content/sigtrans/8/382/ra61.full>
<http://stke.sciencemag.org/content/sigtrans/8/382/pc15.full>
<http://science.sciencemag.org/content/sci/339/6117/279.full>
<http://science.sciencemag.org/content/sci/348/6239/1094.full>
<http://science.sciencemag.org/content/sci/341/6146/1237905.full>
<http://science.sciencemag.org/content/sci/341/6146/626.full>
<http://stke.sciencemag.org/content/sigtrans/8/404/ra121.full>
<http://science.sciencemag.org/content/sci/302/5646/890.full>
<http://stke.sciencemag.org/content/sigtrans/9/436/ec159.abstract>
<http://science.sciencemag.org/content/sci/353/6296/300.full>
<http://stke.sciencemag.org/content/sigtrans/9/437/ec167.abstract>

References This article cites 84 articles, 21 of which you can access for free at:
<http://stke.sciencemag.org/content/8/404/ra119#BIBL>

Permissions Obtain information about reproducing this article:
<http://www.sciencemag.org/about/permissions.dtl>

3.6 Tumor hypoxia causes DNA hypermethylation by reducing TET activity

Bernard Thienpont*, Jessica Steinbacher*, Hui Zhao*, Flora D'Anna*, Anna Kuchnio, Mathew Coleman, Bart Ghesquière, Laurien Van Dyck, Athanasios Ploumakis, Bram Boeckx, Luc Schoonjans, Els Hermans, Frederic Amant, Vessela N. Kristensen, Kian Peng Koh, Massimiliano Mazzone, Thomas Carell, Peter Carmeliet, Diether Lambrechts

(* geteilte Erstautorenschaft)

Prolog

Eine Hypermethylierung der Promotoren von Tumor Suppressor Genen führt zu einem gesteigerten Wachstum von Krebszellen. Wie es dazu kommt ist allerdings noch wenig verstanden. In diesem Manuskript wird gezeigt, dass die Mangelversorgung an Sauerstoff in Tumorzellen die Aktivität der Sauerstoff abhängigen Tet Enzyme reduziert. Dies tritt unabhängig von Hypoxie bedingten Veränderungen in der Tet Genexpression, dem Basalmetabolismus, der HIF Aktivität oder nukleären reaktiven Sauerstoffspezies auf und ist lediglich von der Knappheit an Sauerstoff beeinflusst. Die durch Hypoxie reduzierte Aktivität der Tet Enzyme verstärkt die Hypermethylierung an Promotoren *in vitro*. Auch in Partienten finden sich wesentlich mehr methylierten Promotoren in hypoxischen als in normoxischen Tumoren. Die betroffenen Gene sind häufig an der DNA Reparatur, der Regulation des Zellzyklus, an der Angiogenese und an der Metastasierung beteiligt. Etwa 49% der tumorassoziierten Hypermethylierung ist der Hypoxie zuzuschreiben. Somit scheint die Hypoxie in Tumoren als neuartiger Regulator zu fungieren, welcher der DNA Methylierung unterliegt.

Autorenbeitrag

Der Verdau der genomischen DNA, die Entwicklung einer geeigneten Messmethode sowie die massenspektrometrische Quantifizierung und die Auswertung der Daten wurde von mir vorgenommen. Für eine weitere Auflistung des Beitrags der übrigen

Autoren siehe Manuskript.

Lizenz

Kopie der *submitted* Publikation mit Erlaubnis des Verlags.

Tumour hypoxia causes DNA hypermethylation by reducing TET activity

Bernard Thienpont^{1,2*}, Jessica Steinbacher^{3*}, Hui Zhao^{1,2*}, Flora D'Anna^{1,2*}, Anna Kuchnio^{1,4}, Athanasios Ploumakis⁵, Bart Ghesquière¹, Laurien Van Dyck^{1,2}, Bram Boeckx^{1,2}, Luc Schoonjans^{1,4}, Els Hermans⁶, Frederic Amant⁶, Vessela N. Kristensen^{7,8}, Kian Peng Koh⁹, Massimiliano Mazzone^{1,10}, Mathew L. Coleman⁵, Thomas Carell³, Peter Carmeliet^{1,4} & Diether Lambrechts^{1,2}

Hypermethylation of the promoters of tumour suppressor genes represses transcription of these genes, conferring growth advantages to cancer cells. How these changes arise is poorly understood. Here we show that the activity of oxygen-dependent ten-eleven translocation (TET) enzymes is reduced by tumour hypoxia in human and mouse cells. TET enzymes catalyse DNA demethylation through 5-methylcytosine oxidation. This reduction in activity occurs independently of hypoxia-associated alterations in TET expression, proliferation, metabolism, hypoxia-inducible factor activity or reactive oxygen species, and depends directly on oxygen shortage. Hypoxia-induced loss of TET activity increases hypermethylation at gene promoters *in vitro*. In patients, tumour suppressor gene promoters are markedly more methylated in hypoxic tumour tissue, independent of proliferation, stromal cell infiltration and tumour characteristics. Our data suggest that up to half of hypermethylation events are due to hypoxia, with these events conferring a selective advantage. Accordingly, increased hypoxia in mouse breast tumours increases hypermethylation, while restoration of tumour oxygenation abrogates this effect. Tumour hypoxia therefore acts as a novel regulator of DNA methylation.

Although the mutagenic processes underlying oncogenesis are well studied, tumours are known to be not only genetically but also epigenetically distinct from their tissue of origin. The most extensively documented examples of oncogenic epigenetic changes are those to DNA methylation, but the underlying mechanisms are poorly understood¹.

In tumours, changes in DNA methylation involve both global hypomethylation and the local hypermethylation of CpG-rich gene promoters¹. Hypermethylation frequently affects tumour suppressor genes (TSGs), downregulating their expression and thus contributing to oncogenesis. It remains unclear how methylation changes arise, but an instructive model suggests that genetic changes are a prerequisite for methylation changes²; *BRAF* mutations, for instance, lead to hypermethylation in colorectal tumours³. This is problematic as, while pervasive, hypermethylation of TSGs can only be explained by somatic mutations in a fraction of tumours. Notably, extensive hypermethylation can be seen in ependymomas completely devoid of somatic mutations⁴.

In contrast to DNA methylation mechanisms, those of demethylation have remained elusive until recently, when TET methylcytosine dioxygenases (TET1, TET2 and TET3) were shown to oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC)⁵. 5-Hydroxymethylcytosine and its further-oxidized derivatives are subsequently replaced with an unmodified cytosine by base-excision repair to achieve demethylation⁶. Reduced 5mC oxidation due to decreased TET activity thus increases levels of DNA methylation. Mutations suppressing TET activity are often found in myeloid leukaemia and glioblastoma^{6–9}, but less frequently in other cancer types. By contrast, 5hmC loss is pervasive in tumours and even proposed as a cancer hallmark¹⁰. As with hypermethylation, somatic mutations explain the loss of 5hmC in only a fraction of tumours and it remains unclear which other factors trigger this loss².

Notably, like hypoxia-inducible factor (HIF)-prolyl-hydroxylase domain proteins (PHDs), TET enzymes are Fe²⁺- and α -ketoglutarate-dependent dioxygenases¹¹. PHDs are oxygen-sensitive, acting as oxygen sensors. Under normoxic conditions, they hydroxylate the HIF transcription factors, targeting them for proteasomal degradation, whereas under hypoxia they do not, leading to HIF stabilization and hypoxia response activation¹². Expanding tumours continuously become disconnected from their vascular supply, resulting in vicious cycles of hypoxia, HIF activation and tumour vessel formation¹³. Consequently, hypoxia pervades in solid tumours. Oxygen levels range from 5% to anoxia and around one-third of tumour areas contain less than 0.5% oxygen¹⁴. Although DNA hypermethylation and hypoxia are well-recognized cancer hallmarks, the effect of hypoxia on TET hydroxylase activity and subsequent DNA de-methylation has not been assessed. We therefore set out to investigate whether a hypoxic micro-environment decreases TET hydroxylase activity in tumours, leading to an accumulation of 5mC and acquisition of hypermethylation.

Effect of hypoxia on DNA hydroxymethylation

To assess whether hypoxia affects TET activity, we exposed ten human and five murine cell lines with detectable 5hmC levels to 21% O₂ (normoxic) or 0.5% O₂ (hypoxic, commonly observed in tumours¹⁴) for 24 h. Hypoxia induction was verified and DNA was extracted and profiled for nucleotide composition using liquid chromatography–mass spectrometry (LC–MS). We observed 5hmC loss in eleven cell lines, including eight cancer cell lines (Fig. 1a). However, this did not translate into global 5mC increases (Extended Data Fig. 1), presumably because 5mC is more abundant and is not targeted by TETs at many sites¹⁵. The effect of hypoxia was concentration- and time-dependent:

¹Vesalius Research Center, VIB, 3000 Leuven, Belgium. ²Laboratory of Translational Genetics, Department of Oncology, KU Leuven, 3000 Leuven, Belgium. ³Center for Integrative Protein Science, Department für Chemie und Pharmazie, Ludwig-Maximilians-Universität, 81377 München, Germany. ⁴Laboratory of Angiogenesis and Vascular Metabolism, Department of Oncology, KU Leuven, 3000 Leuven, Belgium. ⁵Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B15 2TT, UK. ⁶Gynecologic Oncology, University Hospitals Leuven, Department of Oncology, KU Leuven, 3000 Leuven, Belgium. ⁷Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, N-0310 Oslo, Norway. ⁸Department of Clinical Molecular Biology (EpiGen), Akershus University Hospital and Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Postboks 1171, Blindern 0318 Oslo, Norway. ⁹Department of Development and Regeneration, and Stem Cell Institute Leuven, KU Leuven, 3000 Leuven, Belgium. ¹⁰Laboratory of Molecular Oncology and Angiogenesis, Department of Oncology, KU Leuven, 3000 Leuven, Belgium.

*These authors contributed equally to this work.

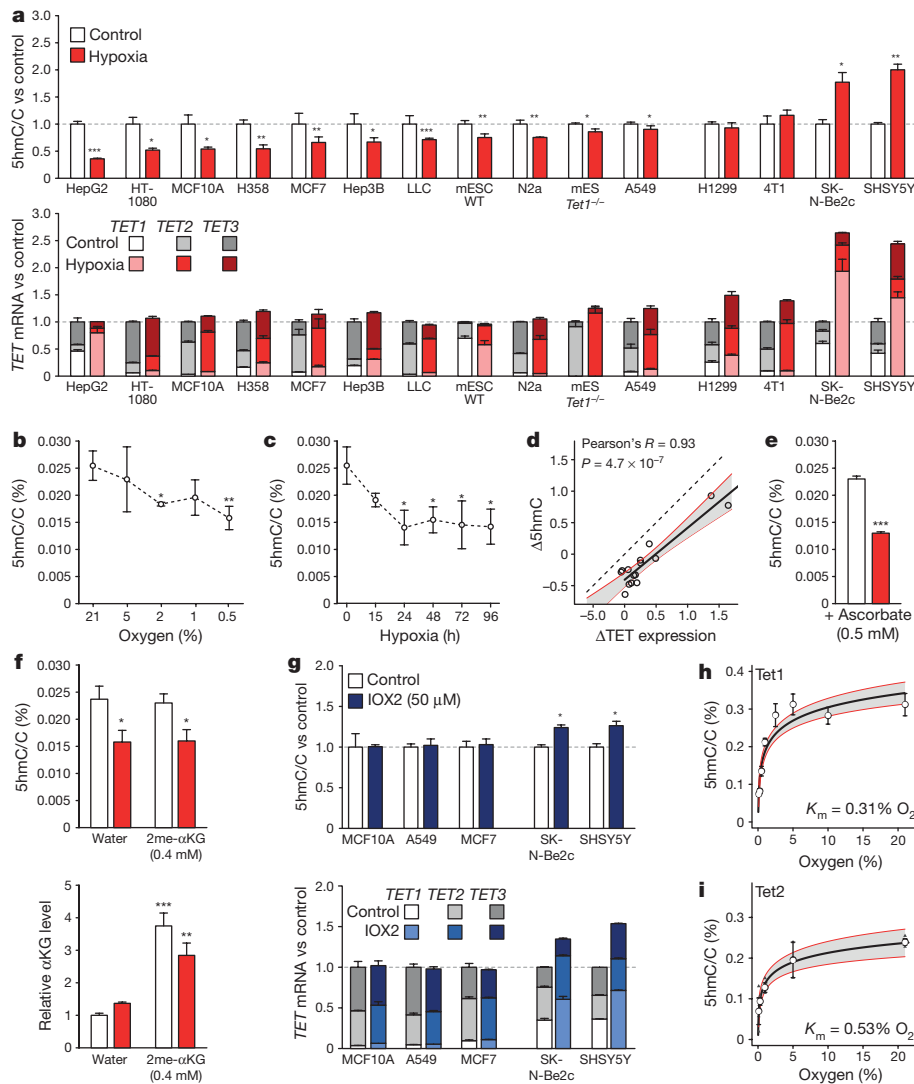


Figure 1 | Effect of hypoxia on 5hmC *in vitro*. **a**, Levels of 5hmC (top), and overall *TET* expression (bottom) in cell lines grown for 24 h under 21% or 0.5% O_2 . RNA expression is expressed relative to the combined estimated level of all 3 *TET* paralogues under 21% O_2 . **b**, **c**, 5hmC/C levels in MCF7 cells exposed to different O_2 levels for 24 h (**b**), or 0.5% O_2 for indicated times (**c**). **d**, Correlation of changes in overall *TET* expression and 5hmC upon hypoxia. Each circle represents a cell line, the full line the correlation. **e**, **f**, Levels of 5hmC (**e**) and α -ketoglutarate (α KG) (**f**) in

MCF7 cells grown with ascorbate (**e**), water or dimethyl- α -ketoglutarate (2me- α KG) (**f**) under 21% (white) or 0.5% O_2 . α -ketoglutarate changes are relative to matching water controls. **g**, As in **a**, but for cells exposed to IOX2. **h**, **i**, Michaelis-Menten curve of Tet1 (**h**) and Tet2 (**i**, $n = 3$) for O_2 . K_m denotes Michaelis constant. Error bars denote s.e.m., grey areas: 95% confidence interval, $n = 5$ replicates for **a**–**h**, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ by *t*-test (**b**, **c**, **e**) or analysis of variance (ANOVA) with post-hoc Tukey HSD test (**f**).

a dose–response revealed loss of 5hmC at oxygen levels at-or-below 2%, and a 20% and 40% reduction, respectively, after 15 h and >24 h of hypoxia (Fig. 1b, c). Loss of 5hmC was not due to increased 5hmC oxidation to 5fC¹⁶, as hypoxia also decreased 5fC levels in embryonic stem (ES) cells (Extended Data Fig. 1).

In some cell lines, levels of 5hmC failed to decrease under hypoxia. 5hmC levels were unaffected in cell lines H1299 and 4T1, and even increased in SHSY5Y and SK-N-Be2c neuroblastoma cells, as reported previously¹⁷ (Fig. 1a). When profiling *TET* expression, neuroblastoma cells displayed potent hypoxia-dependent induction of *TET1* and *TET2*. Cell lines H1299 and 4T1 exhibited intermediate increases in expression levels, whereas all other cell lines showed no, or modest, increases of some *TET* paralogues (Fig. 1a). *Tet* gene expression changes were confirmed at the protein level in mouse cell lines, and HIF1 β -chromatin immune precipitation followed by sequencing (ChIP-seq) further confirmed that HIF binds near the promoters of upregulated *Tet* genes, but not near those that are unaltered (Extended Data Fig. 2a, b), in keeping with the cell-type specificity of the hypoxia response¹². Notably, no cell line showed

decreased *Tet* expression, indicating that 5hmC loss is not due to reduced *Tet* expression.

Since hypoxia affects *TET* paralogue expression differently in different cell lines we correlated hypoxia-associated changes in overall *TET* expression (the combined abundances of *TET1*, *TET2* and *TET3*) with changes in 5hmC levels. Hypoxia reduced 5hmC levels by an average of 44% ($P = 0.0097$) in each cell line (Fig. 1d), independently of *TET* expression changes. Nevertheless, changes in *TET* expression also affected 5hmC levels. This was confirmed by short interfering RNA (siRNA) knockdown of *TET2*, which constitutes around 60% of all *TET* expression in MCF7 cells. This reduced 5hmC levels by around 60% (Extended Data Fig. 2c). Similarly, *Tet1* knockout mouse ES cells (*Tet1*^{−/−}) displayed lower 5hmC levels than wild-type ES cells, in which *Tet1* is the predominantly expressed *Tet* paralogue under both normoxic and hypoxic conditions (Fig. 1a, Extended Data Fig. 2d).

Post-hypoxic 5hmC levels therefore appear to be determined by altered oxygen availability and by changes in *TET* abundance. This explains why cell lines that do not upregulate *TET* expression in response to hypoxia display 5hmC loss, whereas cell lines that strongly

upregulate *TET* compensate for this, resulting in equal or increased 5hmC levels.

Hypoxia directly affects DNA hydroxymethylation

Aside from gene expression, *TET* activity is affected by a variety of cellular processes, including changes in levels of reactive oxygen species (ROS), Krebs cycle metabolites and proliferation^{7,11,17,18}. Since such changes might also occur secondary to hypoxia, we investigated whether they underlie 5hmC reductions in hypoxia.

First, we assessed whether ROS could affect *TET*s in the nucleus through inactivation of Fe^{2+} in their catalytic domain. Although hypoxia increased overall ROS levels, no increase in nuclear ROS was detected either by a nucleus-specific ROS probe or through 8-oxoguanine (8-oxoG) quantification (Extended Data Fig. 3a–f). Moreover, ascorbate supplementation to counteract ROS increases¹⁹ failed to rescue 5hmC loss (Fig. 1e).

Second, because changes in metabolites such as succinate and fumarate compete with *TET* for its cofactor α -ketoglutarate⁷, we investigated whether this was relevant. The concentration of these metabolites, however, was not increased in hypoxic MCF10A or embryonic stem (ES) cells, and only 3–4-fold in MCF7 cells (Extended Data Fig. 3g–i). Levels of the onco-metabolite 2-hydroxyglutarate were also increased in hypoxic MCF7 and MCF10A cells, but were still only around 5–10% of α KG (Extended Data Fig. 3h, j), and therefore unlikely to affect *TET* activity, as affinity of these competing metabolites for hydroxylases is lower or similar to α KG^{7,20}. Culturing MCF7 cells in glutamine-free medium to decrease the concentration of these metabolites did not alter 5hmC levels (Extended Data Fig. 3k). Similarly, exogenous addition of cell-permeable α KG under hypoxia to counteract putative competing metabolites did not rescue the 5hmC loss (Fig. 1f). This therefore precluded metabolite competition from causing hypoxia-associated 5hmC loss.

Third, increases in cell proliferation have also been linked to 5hmC loss²¹. However, cell growth was unaffected or decreased upon exposure to hypoxia in all cell lines tested, indicating that increased proliferation does not underlie 5hmC reduction (Extended Data Fig. 3l).

Fourth, to exclude any potential cellular changes caused by HIF activation, we pharmacologically activated the hypoxia response program by exposing five cell lines grown in atmospheric conditions to IOX2, a small molecule inhibitor with high specificity for PHDs²² (Extended Data Fig. 3m). Cell lines not characterized by hypoxia-induced *TET*-expression changes (MCF10A, A549 and MCF7) showed no change in 5hmC levels under IOX2, while those characterized by *TET* upregulation (SK-N-BE2c and SHSY5Y) showed an increase in 5hmC (Fig. 1g). Thus, after IOX2 exposure, changes in 5hmC levels mirrored changes in *TET* transcription. We also prepared nuclear protein extracts from MCF7 cells grown under hypoxic and atmospheric conditions, and then compared their 5mC oxidative capacities at the same oxygen tension *in vitro*. These, however, were identical (Extended Data Fig. 3n). Loss of 5hmC was therefore not due to activation of the hypoxia response program.

Finally, we assessed the effect of varying oxygen concentrations on the activity of recombinant purified Tet1 or Tet2, by measuring conversion of 5mC to 5hmC on double-stranded genomic DNA. We observed a dose-dependent reduction in 5hmC production with decreasing concentrations. Importantly, under the hypoxic conditions applied in this study (0.5% O_2), Tet1 and Tet2 activity were reduced by $45\% \pm 7$ and $52\% \pm 8$ (mean \pm s.e.m., $P = 0.01$; Fig. 1h, i).

Together, these data demonstrate that decreased oxygen availability directly diminishes the oxidative activity of *TET*s, independently of changes in HIF activity, competing metabolites, proliferation, nuclear ROS or *TET* expression.

Loci with differential DNA hydroxymethylation

To analyse where in the genome hypoxia reduces 5hmC. DNA from hypoxic and normoxic MCF7 cells was immunoprecipitated using

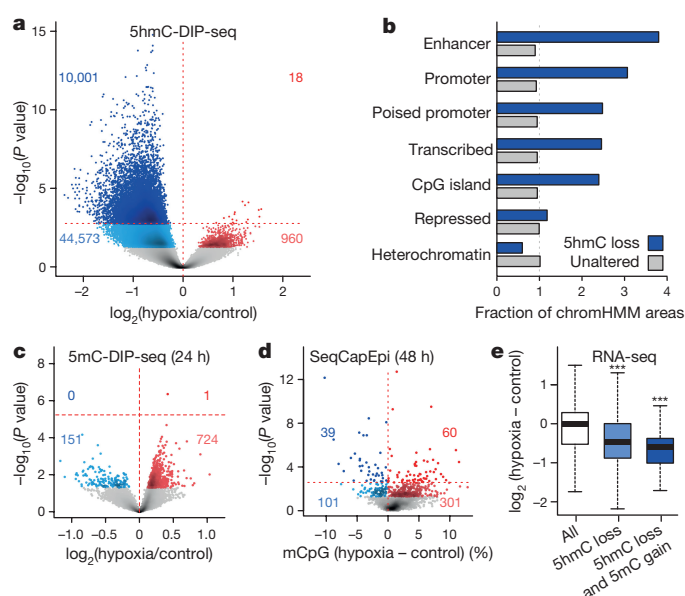


Figure 2 | Genomic profiles of 5(h)mC in MCF7 following hypoxia.

a, Changes in 5hmC at 290,382 peaks detected using 5hmC-DIP-seq. Peaks gaining (red) and losing (blue) 5hmC are highlighted at $P < 0.05$ and 5% FDR adjustment (lighter and darker). **b**, Observed/expected fraction of 5hmC peaks overlapping with chromHMM chromatin states either exhibiting hypoxia-associated 5hmC loss ($n = 10,001$, blue) or not ($n = 280,381$, grey). **c**, **d**, Changes in 5mC after 24 h (**c**) or 48 h (**d**) of 0.5% O_2 , assessed by 5mC-DIP-seq at 10,001 hypohydroxymethylated peaks upon hypoxia (**c**) or by BS-seq at 1,894 regions capture-selected using SeqCapEpi (**d**; see Methods). **e**, Expression changes of genes in hypohydroxymethylated, and both hypohydroxymethylated and hypermethylated peaks. Plots depict 3 (**a**, **e**), 4 (**c**) or 5 (**d**) replicates, *** $P < 0.001$ by negative binomial generalized linear models (**a**, **c**), Fisher's exact test (**d**) or *t*-test (**e**).

antibodies targeting 5mC or 5hmC and subjected to high-throughput sequencing (DIP-seq). We detected 290,382 sites enriched for 5hmC. After hypoxia, 10,001 of the peaks generated for each site exhibited a decrease in 5hmC (5% false discovery rate (FDR)) and only 18 exhibited an increase, thereby confirming global 5hmC loss (Fig. 2a, Supplementary Table 1). Genomic annotation of these peaks using chromHMM²³ revealed they were predominantly found at gene promoters, with some at enhancers and actively transcribed regions, in line with known *TET*-binding sites¹⁵ (Fig. 2b). For example, 5hmC was decreased near transcription start sites of TSGs *NSD1*, *FOXA1* and *CDKN2A* (Extended Data Fig. 4). Analysis of 5mC-DIP signals at these 10,001 regions highlighted that, in 724 out of 875 altered regions, the 5mC content was significantly increased ($P < 0.05$), although only one of these sites survived a 5% FDR correction (Fig. 2c, Supplementary Table 2). Increases in 5mC were thus more subtle than decreases observed for 5hmC.

Several days may be required for 5hmC changes to translate into 5mC changes¹⁹. We therefore cultured cells for 48 h (rather than 24 h) under hypoxia, and used targeted bisulfite-sequencing (BS-seq) to obtain base-resolution quantitation of 5mC at around 85 Mb of promoters and enhancers. Using this approach, we could assess increases in 5mC for 1,894 of the 10,001 regions displaying 5hmC loss. As observed after 5mC-DIP-seq, 301 out of 402 altered sites displayed increased methylation ($P < 0.05$). Similarly, 60 out of 99 altered sites were increased with 5% FDR correction ($P = 2.8 \times 10^{-3}$; Fig. 2d, Supplementary Table 3). ChromHMM annotation revealed that these 60 sites were predominantly in gene promoters and enhancers. To assess the effect of hypermethylation on gene expression, we performed RNA sequencing (RNA-seq) on hypoxic MCF7 cells. Genes depleted in 5hmC and with increased 5mC showed significantly decreased expression upon hypoxia (Fig. 2e; $P = 2.5 \times 10^{-42}$ and 7.4×10^{-4} , respectively, for 3,660

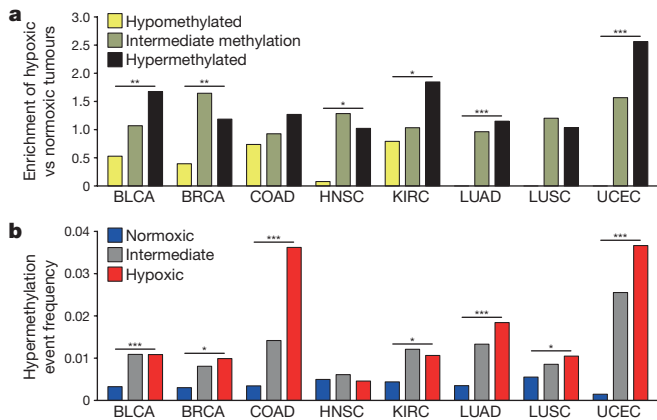


Figure 3 | Effect of hypoxia on hypermethylation in TCGA.

a, Observed and expected number of hypoxic versus normoxic tumours in 3 methylation clusters for 1,000 CpGs hypermethylated in tumour versus normal tissue. **b**, Percentage of hypermethylation events in promoters of frequently hypermethylation genes. $n = 3,141$ tumours, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ by Cochran–Armitage (**a**), generalized linear model per tumour type corrected for co-variants (Supplementary Table 8) (**b**). BLCA, bladder carcinoma; BRCA, breast carcinoma; COAD, colorectal adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; UCEC, uterine corpus endometrial carcinoma.

genes with 5hmC loss and 55 genes with both 5hmC loss and 5mC gain; Supplementary Table 4). Reduced TET activity therefore leads to an accumulation of 5mC, decreasing the expression of associated genes.

Hypermethylation events in hypoxic tumours

We next assessed whether 5hmC loss and concomitant 5mC gain also occur *in vivo*. We focused on gene promoters as they are more frequently affected upon hypoxia, and directly linked to gene expression. As cancer cells go through multiple rounds of sustained hypoxia¹⁴, we proposed that there would be an increase in 5mC, as it would provide a selective advantage for cancer cells, similar to somatic mutations. We therefore assessed 5hmC levels in three patient-derived tumour xenografts, in which we marked hypoxic areas with pimonidazole (Extended Data Fig. 5a). Immunofluorescence analysis revealed decreased 5hmC in hypoxic areas, linking tumour hypoxia to 5hmC loss *in vivo*.

To assess whether hypoxia-associated hypermethylation contributes to the oncogenic process, we analysed tumours profiled in the pan-cancer study of The Cancer Genome Atlas (TCGA)²⁴. We selected 8 solid tumour types (3,141 tumours) for which both DNA methylation (450k array) and gene expression (RNA-seq) data were available for >100 samples, and classified each as hypoxic, normoxic or intermediate using an established gene signature²⁵ (Extended Data Fig. 5b). Next, we analysed tumour-associated DNA hypermethylation in each tumour type by performing unsupervised clustering of 1,000 CpGs that displayed the strongest hypermethylation in tumour versus normal tissue (Extended Data Fig. 5c). In the first three clusters (displaying low, intermediate and high average hypermethylation), we analysed the enrichment of hypoxic tumours. For all eight tumour types, hypoxic tumours predominated in the hypermethylated cluster and normoxic tumours in the hypomethylated cluster (Fig. 3a; $P = 2 \times 10^{-4}$), suggesting that hypoxia leads to increased methylation in tumours.

Whereas the prior analysis identified uniform increases in methylation based on average changes, it poorly captured exceptional increases in hypermethylation known to occur in a subset of tumours^{1,26}. We therefore also modelled tumour hypermethylation by annotating increases in CpG methylation at gene promoters using a stringent threshold (Bonferroni-corrected $P < 0.05$) as hypermethylation events. In each tumour type, the promoters of 187 ± 38 out of 29,649 genes frequently displayed hypermethylation events (Supplementary Table 5).

Notably, hypoxic tumours had on average 4.8-fold more hypermethylation events in these genes than normoxic tumours (Fig. 3b; $P = 4.1 \times 10^{-13}$). These events were functional, reducing gene expression in tumours carrying these hypermethylation events (Extended Data Fig. 5d). They primarily affected promoters with high or intermediate CpG content, in line with TET target preference (Extended Data Fig. 5e)¹⁵. Furthermore, they were not restricted to a small subset: $77\% \pm 6.5$, $49\% \pm 9.3$ or $39\% \pm 9.1$ of hypoxic tumours were affected by ≥ 1 , ≥ 10 or ≥ 20 hypermethylation events, respectively. Considering hypermethylation frequency in normoxic tumours as baseline, up to 48% of hypermethylation events were hypoxia-related.

As hypermethylation can also be genetically encoded, mutations in some genes correlated positively with hypermethylation (for example, *IDH1*, *TET1*, *TET3* and *BRAF*; Supplementary Table 6). Importantly, hypoxia predicted hypermethylation independent of mutation status ($P = 6.1 \times 10^{-12}$). Mutations inhibiting TET activity were infrequent (approximately 1.8% of tumours), indicating that hypermethylation is not genetically encoded in most tumours. *TET*-mutant tumours were also not more hypoxic, suggesting that hypoxia induces hypermethylation, and not vice versa (Extended Data Fig. 5f). Hypoxia-associated hypermethylation events occurred independently of other tumour characteristics such as tumour cell percentage, immune cell infiltration, tumour size, proliferation or metastasis ($P = 4 \times 10^{-13}$), and were significant in seven out of eight tumour types (Supplementary Tables 7, 8). In line with an earlier report²¹, high proliferation was the only other variable significantly predicting hypermethylation ($P = 5.3 \times 10^{-10}$), although only in four of eight tumour types (Extended Data Fig. 5g, h). Using multiple regression, we estimated the contribution of tumour characteristics to hypermethylation variance. On the basis of partial correlation coefficients, proliferation predicted $12.1\% \pm 4.1$, and hypoxia $33.3\% \pm 5.7$, of hypermethylation events explained by the model (Extended Data Fig. 5i).

Given the increase in hypermethylation events in hypoxic tumours, we next selected genes with more hypermethylation events in hypoxic versus normoxic tumours (5% FDR). This revealed 263 ± 94 genes per tumour type, with $9.0\% \pm 1.6$ being shared between any two types (Supplementary Table 9). Ontology analysis of hypermethylated genes revealed that they had biological processes in common such as cell cycle arrest, DNA repair and apoptosis. Hypermethylation was also observed in genes involved in suppressing glycolysis, angiogenesis and metastasis, consistent with tumour hypoxia inducing these processes (Extended Data Fig. 6a–c).

Reduced TET activity underlies hypermethylation

We used three strategies to confirm the role of TET activity in hypoxia-associated hypermethylation. First, we correlated *TET* expression with hypermethylation events, correcting for hypoxia and proliferation. *TET2* and *TET3* expression inversely correlated with hypermethylation ($P = 0.046$ and 0.0028 , Extended Data Fig. 7a), as did hypoxia and proliferation ($P < 1.2 \times 10^{-13}$ for both). Similar to our *in vitro* observations, this implicates reduced TET activity in hypermethylation.

Second, we assessed the overlap of hypermethylation events induced by hypoxia and *IDH1*^{R132} mutations⁸ in 63 glioblastomas. Among wild-type *IDH1* glioblastomas, hypermethylation frequency was 3.4-fold higher in hypoxic tumours (Fig. 4a, Extended Data Fig. 7b). As expected, *IDH1*^{R132} tumours were hypermethylated, albeit 3.9-fold more so than hypoxic tumours (Fig. 4a). This indicates that TET enzymes, fully inactivated in *IDH*-mutant tumours⁹, were only partially inactivated in hypoxia, similar to our *in vitro* observations. Of 228 genes frequently hypermethylated in glioblastomas, those in the hypoxic and *IDH*-mutant subgroups displayed a 58% overlap ($P < 10^{-16}$; Fig. 4b) and reduced expression (Extended Data Fig. 7c), indicating that loss of TET activity affects the same genes, regardless of the underlying trigger.

Finally, to link hypoxia-associated hypermethylation to 5hmC loss, we profiled 24 non-small-cell lung tumours for 5mC and 5hmC using 450k arrays (Extended Data Fig. 7d). This revealed a generalized loss

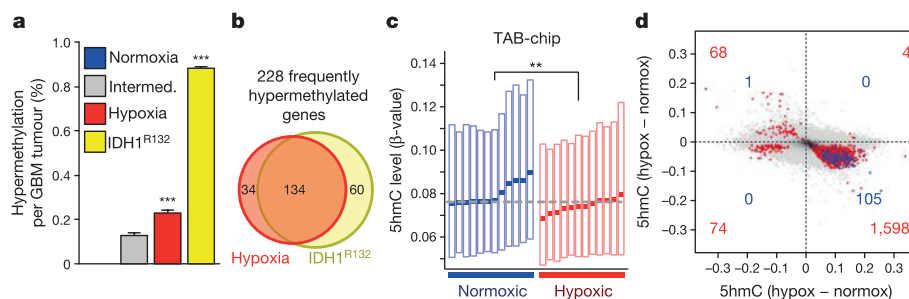


Figure 4 | Effect of hypoxia on TET activity in human tumours.

a, Hypermethylation in 19 normoxic (blue), 21 intermediate (grey), 17 hypoxic (red) and 6 IDH1^{R132}-mutated (yellow) glioblastomas. **b**, Overlap between genes hypermethylated in hypoxic versus IDH1^{R132}-mutated glioblastomas. **c**, 5hmC measured across 485,000 CpGs in 12

normoxic versus 12 hypoxic non-small-cell lung tumours. **d**, Changes in 5(h)mC for unaltered CpGs (grey), and CpGs altered in both 5mC and 5hmC (25% FDR, blue; $P < 0.01$, red). *** $P < 0.001$ by Fisher's exact (**a**), ** $P < 0.01$ by t -test (**c**).

of 5hmC in hypoxic tumours ($-7.1\% \pm 1.1$; $P = 3.7 \times 10^{-3}$; Fig. 4c). Individual probes also mostly displayed 5hmC loss and 5mC gain in hypoxic tumours (96.7% and 65.4% of probes altered, respectively, $P < 0.01$; Supplementary Table 10). Of all probes displaying 5mC gain, most (87%) also displayed 5hmC loss, and of probes altered both in 5hmC and 5mC ($P < 0.01$), 92% showed 5hmC loss and 5mC gain (Fig. 4d; $P < 10^{-16}$). This directly implicates hypoxia-induced loss of 5hmC in the hypermethylation of hypoxic tumours.

Rescue of hypoxia-induced hypermethylation

To manipulate tumour oxygenation and confirm its effect on hypermethylation, we used mice expressing the polyomavirus middle T-antigen under the mouse mammary tumour virus promoter (MMTV-PyMT). These mice spontaneously develop breast tumours, with hypoxic areas emerging from 7 weeks onwards, encompassing approximately 20% of the tumour at 16 weeks²⁷. Hypoxic areas in these tumours were also depleted in 5hmC (Fig. 5a, b).

We monitored hypermethylation changes by targeted BS-seq of TSG promoters commonly inactivated in cancer²⁸. Hypoxic human breast tumours displayed a specific increase in hypermethylation at these TSG promoters, whereas no effect was observed for oncogenes (Extended Data Fig. 8a). In line with the age-associated increase in tumour hypoxia²⁷, hypermethylation events also increased markedly with age (and tumour size), but not in normal mammary glands (Extended Data Fig. 8b–d). Importantly, >95% of cells in these tumours were PyMT-positive, whereas cell proliferation and immune cell infiltration were comparable between hypoxic and normoxic areas (Extended Data Fig. 8e–g). Hypermethylation changes are therefore unlikely to be a result of changes in proliferation or cellular heterogeneity.

To test whether reduced tumour oxygenation increases hypermethylation, 9-week-old MMTV-PyMT mice were hydrodynamically injected with a soluble Flk1 (sFlk1)-expressing plasmid. After 3 weeks, this caused tumour vessel pruning and hypoxia (Extended Data Fig. 9a–d). Shallow whole-genome sequencing for 5hmC (TET-assisted bisulfite sequencing; TAB-seq) revealed a global loss of 5hmC after sFlk1 overexpression ($-12.4\% \pm 3.5$, $P = 0.040$), occurring predominantly at gene-dense regions and affecting the entire gene (Fig. 5c, Extended Data Fig. 9e), consistent with previously described 5hmC distributions¹⁵. Moreover, targeted BS-seq revealed an exacerbated hypermethylation phenotype after sFlk1 overexpression at 12 weeks in TSGs but not oncogenes (10 out of 15 TSGs contained ≥ 1 hypermethylation event; $P = 0.010$, Fig. 5d, Extended Data Fig. 9f). Tumour growth and the expression of proliferation markers, *Tet* paralogues and the immune cell marker CD45 were unaffected by sFlk1 overexpression, indicating that hypermethylation occurs independently (Extended Data Fig. 9g–j).

To rescue this effect, we normalized the tumour vasculature by intercrossing a heterozygous *Phd2* (also known as *EglN1*) loss-of-function allele with the PyMT transgene. This significantly reduced tumour hypoxia at 16 weeks²⁷ (Extended Data Fig. 9k). TAB-seq revealed a 5hmC gain ($+25.3\% \pm 4.7$, $P = 0.0098$) occurring primarily

at gene-dense regions and affecting the entire gene (Fig. 5c, Extended Data Fig. 9l). Notably, BS-seq revealed that, although 8 out of 15 TSGs displayed ≥ 1 hypermethylation event in *Phd2*^{+/+} tumours, no hypermethylation was observed in *Phd2*^{-/-} tumours ($P = 2.6 \times 10^{-7}$, Fig. 5e). Again, oncogenes were unaffected (Extended Data Fig. 9m). Effects were independent of *Phd2* haploinsufficiency in tumour cells, as similar effects were observed in PyMT mice having endothelial-cell-specific *Phd2* haploinsufficiency²⁷ (Extended Data Fig. 9n, o). As in the sFlk1

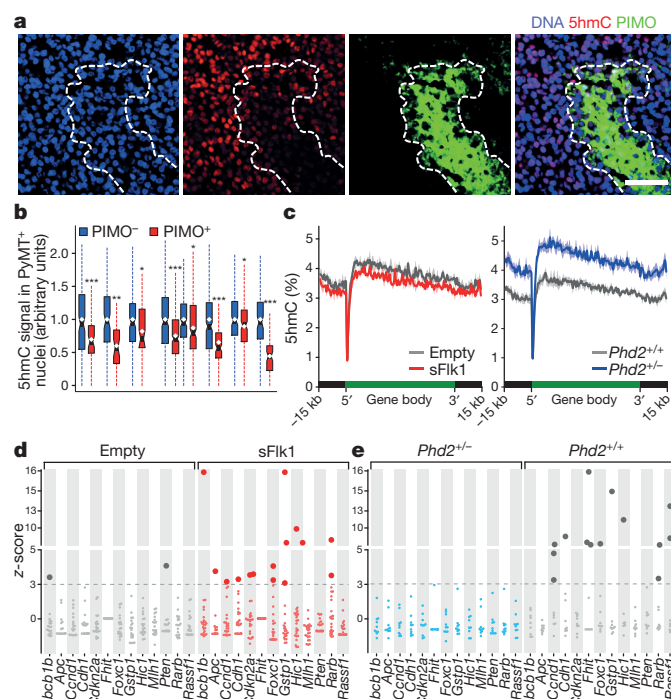


Figure 5 | Effect of vessel pruning and normalization on 5hmC and TSG hypermethylation.

a, **b**, Immunofluorescence of breast tumours in transgenic (MMTV-PyMT) mice. **a**, Representative image. Scale bar, 50 μ m. **b**, Box plot of 5hmC signal in >150 PyMT-positive nuclei from eight tumours, stratified for pimonidazole (PIMO) (yes/no) and normalized to PIMO-negative nuclei. **c**, 5hmC levels \pm s.e.m. across a metagene in tumours of 12-week-old mice receiving empty or sFlk1-overexpressing plasmid (left, $n = 3$), or 16-week-old mice with the indicated genotype (right, $n = 3$ for *Phd2*^{+/+}, $n = 4$ for *Phd2*^{-/-}). **d**, **e**, Hypermethylation in tumours developing in 12-week-old mice receiving empty ($n = 19$) or sFlk1-overexpressing plasmid ($n = 24$) 3 weeks earlier (**d**), and in tumours developing in 16-week-old *Phd2*^{+/+} ($n = 10$) and *Phd2*^{-/-} ($n = 9$) mice (**e**). Plotted are z-scores of hypermethylation, relative to normoxic tumours (empty and *Phd2*^{+/+} for **d** and **e**). Dotted line: 5% FDR, darker dots: significant hypermethylation. *Brca1* and *Timp3*: not shown (no hypermethylation event detected). Hypermethylated genes on average had 5.8% (**d**) and 4.7% (**e**) more methylation. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by t -test.

model, increasing tumour oxygenation by *Phd2* haploinsufficiency did not affect tumour growth or the expression of proliferation markers, *Tet* paralogs or CD45 (Extended Data Fig. 9p–u).

Discussion

We show here that tumour hypoxia directly reduces TET activity, causing a 5hmC decrease predominantly at gene promoters and enhancers. Concomitantly, 5mC increases at these sites and, as with certain genetic mutations, provides a substrate for oncogenic selection *in vivo*²⁶. Since hypoxia prevails in tumours, 5mC changes in TSG promoters are frequent, rendering hypoxic tumours hypermethylated at these sites. Hypermethylation events in tumours have long been suspected to occur through selection of random DNA methylation variants²⁹. However, the identification of genetically encoded hypermethylation challenged this stochastic model². By demonstrating that hypoxia drives hypermethylation, we show that genetically-encoded and tumour-microenvironment-driven models of epimutagenesis co-exist. However, since hypoxia is pervasive, the mechanism described here is relevant for most solid tumours. We found that up to 48% of hypermethylation events were hypoxia-related and effects were replicated in all tumour types investigated, independent of mutation- and proliferation-induced hypermethylation. Modest hypoxia (2–5% O₂) did not affect TET activity, indicating that TET enzymes are not physiological oxygen sensors (unlike PHDs) in line with previous reports³⁰. TET activity only becomes limiting under the pathophysiological oxygen concentrations found in tumours¹⁴. Analogous to somatic *TET* haploinsufficiency, this partial reduction in TET activity contributes to oncogenesis. Our findings also suggest intriguing avenues of investigation into other ischaemia-related pathologies.

Our model provides a mechanism for the association between hypoxia and maladaptive oncogenic processes. Genes affected by hypermethylation were not only involved in cell-cycle arrest, DNA repair and apoptosis, but also glycolysis, metastasis and angiogenesis. High doses of angiogenesis inhibitors stimulate metastatic spreading in mouse cancer models (at least in specific settings)³¹, and tumour hypoxia is considered a driver of this behaviour. The mechanism by which hypermethylation accumulates under hypoxia may underlie these escape mechanisms. By contrast, low levels of angiogenic inhibition can induce tumour vessel normalization, and improve oxygenation³². Our observations in normalized PyMT tumours suggest that the therapeutic benefits of vessel normalization such as decreased metastatic burden²⁷, might occur by inhibiting hypoxia-associated hypermethylation. Countering hypermethylation by inhibiting DNA methylation or by normalizing tumour blood supply may therefore prove to be therapeutically beneficial.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June 2015; accepted 5 July 2016.

Published online 17 August 2016.

1. Esteller, M. Epigenetics in cancer. *N. Engl. J. Med.* **358**, 1148–1159 (2008).
2. Struhl, K. Is DNA methylation of tumour suppressor genes epigenetic? *eLife* **3**, e02475 (2014).
3. Weisenberger, D. J. et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with *BRAF* mutation in colorectal cancer. *Nat. Genet.* **38**, 787–793 (2006).
4. Mack, S. C. et al. Epigenomic alterations define lethal CIMP-positive endymomas of infancy. *Nature* **506**, 445–450 (2014).
5. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
6. Shen, L. et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013).
7. Xiao, M. et al. Inhibition of α -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).
8. Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
9. Xu, W. et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).

10. Yang, H. et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**, 663–669 (2013).
11. Ploumaki, A. & Coleman, M. L. OH, the places you'll go! Hydroxylation, gene expression, and cancer. *Mol. Cell* **58**, 729–741 (2015).
12. Schofield, C. J. & Ratcliffe, P. J. Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* **5**, 343–354 (2004).
13. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353–364 (1996).
14. Vaupel, P., Höckel, M. & Mayer, A. Detection and characterization of tumor hypoxia using pO₂ histography. *Antioxid. Redox Signal.* **9**, 1221–1235 (2007).
15. Williams, K. et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
16. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
17. Mariani, C. J. et al. TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma. *Cell Reports* **7**, 1343–1352 (2014).
18. Zhao, B. et al. Redox-active quinones induces genome-wide DNA methylation changes by an iron-mediated and Tet-dependent mechanism. *Nucleic Acids Res.* **42**, 1593–1605 (2014).
19. Blaschke, K. et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).
20. Koivunen, P. et al. Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484–488 (2012).
21. Bachman, M. et al. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).
22. Chowdhury, R. et al. Selective small molecule probes for the hypoxia inducible factor (HIF) prolyl hydroxylases. *ACS Chem. Biol.* **8**, 1488–1496 (2013).
23. Taberlay, P. C., Statham, A. L., Kelly, T. K., Clark, S. J. & Jones, P. A. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res.* **24**, 1421–1432 (2014).
24. Weinstein, J. N. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
25. Buffa, F. M., Harris, A. L., West, C. M. & Miller, C. J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br. J. Cancer* **102**, 428–435 (2010).
26. Feinberg, A. P. & Irizarry, R. A. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl Acad. Sci. USA* **107** (Suppl. 1), 1757–1764 (2010).
27. Kuchnio, A. et al. The cancer cell oxygen sensor PHD2 promotes metastasis via activation of cancer-associated fibroblasts. *Cell Reports* **12**, 992–1005 (2015).
28. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
29. Landan, G. et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012).
30. Laukka, T. et al. Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. *J. Biol. Chem.* **291**, 4256–4265 (2016).
31. Paez-Ribes, M. et al. Antiangiogenic therapy elicits malignant progression of tumors to increased local invasion and distant metastasis. *Cancer Cell* **15**, 220–231, (2009).
32. Heist, R. S. et al. Improved tumor vascularization after anti-VEGF therapy with carboplatin and nab-paclitaxel associates with survival in lung cancer. *Proc. Natl Acad. Sci. USA* **112**, 1547–1552 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Peuteman, T. Van Brussel, J. Serneels and K. Kurz for assistance, C. Chang for NucPE1, G.-L. Xu for Tet-triple knockout ES cells. H.Z. and B.T. hold FWO-F postdoctoral fellowships. This work was supported by funding from the ERC (CHAMELEON 617595 to D.L.; EU-ERC269073 to P.C.; CHAMELEO 334420 to B.T.), from the FWO-F (G065615N, G070615N) to D.L., from the IUAP (P7/03) and the Flemish Government (Methusalem) to P.C., and from the DFG (EXC114 (CIPSM), grants CA275/8-5, GRK2062/1 and SPP1784) to T.C.

Author Contributions B.T. and D.L. conceived and supervised the project, designed experiments and wrote the manuscript. B.T. and F.D.A. performed *in vitro* experiments and analysed data, helped by L.V.D.; M.L.C. and A.P. analysed Tet Michaelis–Menten kinetics; animal models were provided by E.H., F.A. (xenografts), M.M. (sFlk1), A.K. and P.C. (*Phd2*^{+/−}); V.N.K. contributed ideas, L.S. and K.P.K. provided reagents; J.S. quantified nucleotides by LC–MS, supervised by T.C.; B.G. quantified metabolites. H.Z. analysed TCGA tumours; B.T., H.Z. and B.B. performed bioinformatics and statistics.

Author Information Microarray and sequencing data are available at the Gene Expression Omnibus under accession GSE71403. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.T. (bernard.thienpont@vib-kuleuven.be) or D.L. (diether.lambrechts@vib-kuleuven.be).

Reviewer Information *Nature* thanks R. S. Johnson, M. Rehli and Y. Xiong for their contribution to the peer review of this work.

METHODS

Materials. All materials were molecular biology grade. Unless noted otherwise, all were from Sigma.

Analysis of global 5mC and 5hmC levels in cultured cells. MCF7, MCF10A, A549, H1299, SHSY5Y, Hep G2, Hep 3B2, HT-1080, NCI-H358, LLC, Neuro-2a, 4T1 and SK-N-Be2c cell lines were obtained from the American Type Culture Collection and their identity was not further authenticated. These cell lines are not listed in the database of commonly misidentified cell lines maintained by ICLAC. LLC, Neuro-2a, 4T1, Hep G2, HT-1080, Hep 3B2, MCF7 and A549 cells were cultured at 37 °C in DMEM with 10% fetal bovine serum (FBS), 5 ml of 100 U ml⁻¹ penicillin-streptomycin (Life Technologies) and 5 ml of L-glutamine 200 mM. NCI-H358, H1299 and SK-N-Be2c cell lines were cultured at 37 °C in RPMI 1640 Medium with 10% FBS 1% penicillin-streptomycin and 1% L-glutamine. MCF10A cells were cultured at 37 °C in DMEM/F-12 supplemented with 5% horse serum (Life Technologies), 20 ng ml⁻¹ human epidermal growth factor (Preprotec), 0.5 µg ml⁻¹ hydrocortisone, 100 ng ml⁻¹ cholera toxin, 10 µg ml⁻¹ insulin, and 100 U ml⁻¹ penicillin-streptomycin. The SHSY5Y cell line was cultured at 37 °C in DMEM/F-12 supplemented with 10% FBS, 2% penicillin-streptomycin and 1% non-essential amino acids (MEM). Mouse J1 ES cells were cultured feeder-free in fibroblast-conditioned medium. Cell cultures were confirmed to be mycoplasma-free every month.

Cell line treatment conditions. Control cell cultures were grown at atmospheric oxygen concentrations (21%) with 5% CO₂. To render cultures hypoxic, they were incubated in an atmosphere of 0.5% O₂, 5% CO₂ and 94.5% N₂. Where indicated, IOX2 (50 µM), ascorbate (0.5 mM, a dose known to support TET activity¹⁹) or dimethyl- α -ketoglutarate (0.5 mM) was added to fresh culture medium, using an equal volume of the carrier (DMSO) as a control for IOX2. Cells were plated at a density tailored to reach 80–95% confluence at the end of the treatment. Fresh medium was added to the cells just before hypoxia exposure. For glutamine-free culture experiments, dialysed FBS was added to glutamine-free DMEM, and supplemented with glutamine (4 mM) for the control. Mouse J1 ES cells and *Tet1*-gene-trap ES cells were cultured feeder-free in fibroblast-conditioned medium.

DNA extraction. After exposure to the aforementioned stimuli, cultured cells were washed on ice with ice-cold PBS with deferoxamin (PBS-DFO, 200 µM), detached using cell scrapers and collected by centrifugation (400g, 4 °C). Nucleic acids were subsequently extracted using the Wizard Genomic DNA Purification kit (Promega) according to instructions. All buffers were supplemented with DFO (200 µM) and DNA was dissolved in 80 µl PBS-DFO with RNase A (200 U, NEB) and incubated for 10 min at 37 °C. After proteinase K addition (200 units) and incubation for 30 min at 56 °C, DNA was purified using the QIAquick blood and tissue kit (all buffers supplemented with DFO). It was eluted in 100 µl of a 10 mM Tris, 1 mM EDTA solution (pH 8) and stored at –8 °C until further processing.

LC-ESI-MS/MS of DNA to measure 5mC, 5hmC and 8-oxoG levels. To measure the cytosine, 5mC, 5hmC and 8-oxoG content of the DNA samples, three technical replicates were run for each sample. More specifically, 0.5–2 µg DNA in 25 µl H₂O were digested in an aqueous solution (7.5 µl) of 480 µM ZnSO₄, containing 42 U nuclease S1, 5 U Antarctic phosphatase, and specific amounts of labelled internal standards were added and the mixture was incubated at 37 °C for 3 h in a Thermomixer comfort (Eppendorf). After addition of 7.5 µl of 520 µM [Na]₂-EDTA solution containing 0.2 U snake venom phosphodiesterase I, the sample was incubated for another 3 h at 37 °C. The total volume was 40 µl. The sample was then kept at –20 °C until the day of analysis. Samples were then filtered by using an AcroPrep Advance 96-filter plate 0.2 µm Supor (Pall Life Sciences) and then analysed by liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS), which are performed using an Agilent 1290 UHPLC system and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. DNA samples were digested to give a nucleoside mixture and spiked with specific amounts of the corresponding isotopically labelled standards before LC-MS/MS analysis. The nucleosides were analysed in the positive ion selected reaction monitoring mode (SRM). In the positive ion mode, [M + H]⁺ species were measured.

Determination and comparison of nucleoside concentrations. The resulting cytosine, 5mC, 5hmC and 8-oxoG peak areas were normalized using the isotopically labelled standards, and expressed relative to the total cytosine content (that is, C + 5mC + 5hmC). Concentrations were depicted as averages of independent replicates grown on different days, and compared between hypoxia and normoxia (21% O₂), or between control and treated conditions, using a paired Student's *t*-test. No statistical methods were used to predetermine sample size.

RNA extraction, cDNA synthesis and qPCR. For RNA extraction, cell culture medium was removed, TRIzol (Life Technologies) added and processed according to manufacturer's guidelines. Reverse transcription and qPCR were performed using 2 × TaqMan Fast Universal PCR Master Mix (Life Technologies), TaqMan probes and primers (IDT, sequence in Supplementary Table 12). Thermal cycling

and fluorescence detection were done using a LightCycler 480 Real-Time PCR System (Roche). Taqman assay amplification efficiencies were verified using serial cDNA dilutions, and estimated to be >95%.

mRNA concentration analysis and statistics. Cycle threshold (C_t) values were determined for each sample and gene of interest in technical duplicates, and normalized according to the corresponding amplification efficiency. Per sample, *TET* expression was expressed relative to β -2-microglobulin (human) or hypoxanthine phosphoribosyltransferase 1 (*Hprt* mouse) levels by subtraction of their average C_t values. Concentrations were expressed as averages of at least 5 replicates extracted on different days. For Fig. 1a, copy number estimates for *TET1*, *TET2* and *TET3* were expressed for each cell line, relative to the summed copy number estimates of *TET1*, *TET2* and *TET3* under control conditions (21% O₂). Concentrations were compared between hypoxia and normoxia, or between control and treatment conditions using a Student's *t*-test. No statistical methods were used to predetermine sample size.

Hypoxia marker gene induction. To verify further induction of the hypoxia response program, hypoxia marker gene expression was verified. We analysed mRNA levels of genes encoding the E1B 19K/Bcl-2-binding protein Nip3 (*BNIP3*) and fructose-bisphosphate aldolase (*ALDOA*), 2 established hypoxia marker genes³³. Reverse transcriptase-quantitative PCR (RT-qPCR) was performed as described for the *TET* mRNA concentration assays, and differential expression was calculated using the $\Delta\Delta C_t$ method³⁴. We ruled out transcriptional upregulation as the cause of the increase in HIF1 α protein concentrations by assessing *HIF1A* mRNA expression in parallel. mRNA concentrations were expressed relative to normoxic controls (21% O₂). Differences in mRNA concentration were assessed using a Student's *t*-test on 5 or more independent replicates grown on different days.

Western blotting for Hif1 α , Tet1, Tet2 and Tet3. To assess Hif1 α protein stabilization, proteins were extracted from cultured cells as follows: cells were placed on ice, and washed twice with ice-cold PBS. Proteins were extracted with extraction buffer (50 mM Tris HCl, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate and 0.1% SDS) with 1 × protease inhibitor cocktail. Protein concentrations were determined using a bicinchoninic acid protein assay (BCA, Thermo Scientific) following the manufacturer's protocol. An estimated 60 µg protein was loaded per well on a NuPAGE Novex 3–8% Tris-Acetate Protein gel (Life Technologies), separated by electrophoresis and blotted on polyvinylidene fluoride membranes. Membranes were activated with methanol, washed and incubated with antibodies targeting β -actin (4967, Cell Signaling), Tet1 (09-872, Millipore) and Tet3 (61395, Active Motif), at 1:1,000 dilution, targeting Tet2 (124297, Abcam) at 1:250 dilution, and targeting Hif1 α (C-Term) (Cayman Chemical Item 10006421) at 1:3,000 dilution. Secondary antibodies and detection were according to routine laboratory practices. Western blotting was performed on 6 independent replicates grown on different days.

Analysis of HIF1 β target genes using ChIP-seq. To confirm that hypoxia-associated differential expression of *TET* genes is induced by the HIF pathway, we performed HIF1 β ChIP-seq. Because HIF1 β is the obligate binding partner of all three HIF α proteins stabilized and activated upon hypoxia³⁵, HIF1 β ChIP-seq reveals all direct HIF-target genes.

Chromatin immunoprecipitation. Approximately 25 × 10⁶–30 × 10⁶ cells were incubated in hypoxic conditions for 16 h. Cultured cells were subsequently immediately fixed by adding 1% formaldehyde (16% formaldehyde (w/v), Methanol-free, Thermo Scientific) directly to the medium and incubating for 8 min. Fixed cells were incubated with 150 µM of glycine for 5 min to revert cross-links, washed twice with ice-cold PBS 0.5% Triton X-100, scraped and collected by centrifugation (1,000g for 5 min at 4 °C). The pellet was re-suspended in 1,400 µl of RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA pH 8, 1% Triton X-100, 0.5% sodium deoxycholate, 1% SDS, 1% protease inhibitors) and transferred to a new Eppendorf tube. The lysate was homogenized by passing through an insulin syringe, and incubated on ice for 10 min. The chromatin was sonicated for 3 min by using a Branson 250 Digital Sonifier with 0.7 s 'On' and 1.3 s 'Off' pulses at 40% power amplitude, yielding a size of 100 to 500 bp. The sample was kept ice cold at all times during the sonication. The samples were centrifuged (10 min at 16,000g at 4 °C) and the supernatant was transferred in a new Eppendorf tube. The protein concentration was assessed using a BCA assay. Fifty microlitres of shared chromatin was used as 'input' and 1.4 µg of primary ARNT/HIF-1 β monoclonal antibody (NB100-124, Novus) per 1 mg of protein was added to the remainder of the chromatin, and incubated overnight at 4 °C in a rotator. Pierce Protein A/G Magnetic Beads (Life Technologies) were added to the samples in a volume four times the volume of the primary antibody and incubated at 4 °C for at least 5 h. A/G Magnetic Beads were collected and the samples were washed five times with the washing buffer (50 mM Tris-HCl, 200 mM LiCl, 2 mM EDTA, pH 8, 1% Triton, 0.5% sodium deoxycholate, 0.1% SDS, 1% protease inhibitors), and twice with a 10 mM Tris, 1 mM EDTA (TE) buffer. The A/G magnetic beads were re-suspended

in 50 µl of TE buffer, and 1.5 µl of RNase A (200 units, NEB) were added to the A/G beads samples and to the input, incubated for 10 min at 37 °C. After addition of 1.5 µl of proteinase K (200 U) and overnight incubation at 65 °C, the DNA was purified using 1.8 × volume of Agencourt AMPure XP (Beckman Coulter) according to the manufactory instructions, and then eluted in 15 µl of TE buffer. The input DNA was quantified on NanoDrop.

ChIP-seq, mapping and analysis. In total, 5 µg of input and all of the immunoprecipitated DNA was converted into sequencing libraries using the NEBNext DNA library prep master mix set. A single end of these libraries was sequenced for 50 bases on a HiSeq 2000, mapped using Bowtie and extended for the average insert size (250 bases). ChIP peaks were called by model-based analysis for ChIP-Seq³⁶, with standard settings and using a sequenced input sample as baseline.

Patient-derived xenografted tumours. To assess whether tumour-associated hypoxia reduces 5hmC levels *in vivo*, redundant material from two endometrial tumours and a breast tumour, removed during surgery, was grafted in the interscapular region of nude mice. Informed consent was obtained from the patient, following the ethical approval of the local ethical committee. All animal experiments were approved by the local ethical committee (P098/2014). Each tumour was allowed to grow to 1 cm³, after which it was collected. 10% of this tumour was re-implanted in a nude mouse, and the tumour was propagated for three generations until it was used for this experiment. To mark hypoxic areas, mice were injected with pimonidazole (60 mg kg⁻¹, Hypoxyprobe) i.p. 1 h before killing.

Immunofluorescence staining and analysis. Tumours were collected, fixed in formaldehyde and embedded in paraffin using standard procedures. Paraffin was removed and slides were rehydrated in two xylene baths (5 min), followed by five 3-min ethanol baths at decreasing concentrations (100%, 96%, 70%, 50% and water) and a 3-min TBS (50 mM Tris, 150 mM NaCl, pH 7.6) bath.

The following antibodies were used for immunofluorescence staining: primary antibodies were FITC-conjugated mouse anti-pimonidazole (HP2-100, Hydroxyprobe), rabbit anti-5hmC (39791, Active Motif), rat anti-polyoma middle T (AB15085, Abcam), rat anti-CD31 (557355, BD Biosciences), rat anti-CD45 (553076, BD Biosciences), rabbit anti-Ki67 (AB15580, Abcam) and mouse anti-pan cytokeratin (C2562, Sigma). Secondary antibodies were Alexa Fluor 405-conjugated goat anti-rabbit (A31556, Thermo Fisher), Alexa Fluor 647 conjugated goat anti-rat (A-21247, Life Technologies), peroxidase-conjugated goat anti-FITC (PA1-26804, Pierce), biotinylated goat anti-rat (A10517, Thermo Fisher) and biotinylated goat anti-rabbit (E043201, Dako). Signal amplification was performed using the TSA Fluorescein System (NEL701A001KT, Perkin Elmer) or the TSA Cyanine 5 System (NEL705A001KT, Perkin Elmer).

Different protocols were implemented depending on the epitopes of interest. Staining for the following epitopes was combined: CD45, 5hmC, pimonidazole and DNA; PyMT, 5hmC, pimonidazole and DNA; Ki67, pimonidazole and DNA; CD31 and pimonidazole; and pan-cytokeratin, 5hmC, pimonidazole and DNA.

Antigen retrieval for CD31, CD45 and pan-cytokeratin was done by a 7-min trypsin digestion, for pimonidazole and Ki67 using AgR at 100 °C for 20 min, followed by cooling for 20 min. Slides were washed in TBS for 5 min, permeabilized in 0.5% Triton X-100 in PBS for 20 min. For 5hmC antigen retrieval, slides were denatured in 2 M HCl for 10 min; HCl was neutralized for 2 min in borax, 1% in PBS pH 8.5, and washed twice for 5 min in PBS.

For all slides, endogenous peroxidase activity was quenched using H₂O₂ (0.3% in methanol), followed by three 5-min washes in TBS. Slides were blocked using pre-immune goat serum (X0907, Dako; 20% in TNB; TSA Biotin System kit, Perkin Elmer). Binding of primary antibodies (anti-5hmC, anti-CD45, anti-CD31 and anti-pan cytokeratin or FITC-conjugated anti-pimonidazole; all 1:100 in TNB) was allowed to proceed overnight. Slides were washed 3 times in TNT (0.5% Triton-X100 in TBS) for 5 min, after which the following secondary antibodies (all 1:100 in TNB with 10% pre-immune sheep serum) were allowed to bind for 45 min: sheep anti-FITC-PO (for pimonidazole), goat anti-rabbit-Alexa Fluor 405 (for 5hmC), goat anti-rat-Alexa Fluor 647 (for CD45), and biotinylated goat anti-mouse (for pan-cytokeratin). Slides were washed three times for 5 min in TNT, after which signal amplification was performed for 8 min using Fluorescein Tyramide (1:50 in amplification diluent).

Slides stained for pimonidazole that required co-staining for Ki67 or PyMT, or slides stained for pan-cytokeratin that required co-staining for pimonidazole were subjected to a second indirect staining for the latter epitopes. After 5 min of TNT and 5 min of TBS, slides were quenched again for peroxidase activity using H₂O₂ and blocked using pre-immune goat serum, prior to a second overnight round of primary antibody binding (anti-Ki67, FITC-anti-pimonidazole or anti-PyMT, all 1/100). The next day, three 5-min washes with TNT were followed by a 1-h incubation with a biotinylated goat anti-rabbit antibody (for Ki67) or goat anti-rat (for PyMT), another three 5-min washes with TNT, a 30-min incubation with peroxidase conjugated to streptavidin (for Ki67 and PyMT) or to anti-FITC (for pimonidazole), another three 5-min washes with TNT and signal amplification

for 8 min using, for pimonidazole, Fluorescein Tyramide and for others Cyanine 5 Tyramide (1:50 in amplification diluent). Slides were then stained with propidium iodide with RNase (550825; BD biosciences) for 15 min, washed for 5 min in PBS and mounted with Prolong Gold (Life Technologies).

Slides were imaged on a Nikon A1R Eclipse Ti confocal microscope. Three to five sections per slide were imaged, and processed using ImageJ. Nuclei were identified using the propidium iodide signal and nuclear signal intensities for Fluorescein and Cy3 (pimonidazole and 5hmC) measured. Analyses were exclusively performed on slide regions showing a regular density and shape of nuclei, in order to avoid inclusion of acellular or necrotic areas. The pimonidazole signal will also not stain necrotic/acellular areas³⁷, and was used to stratify viable cell nuclei into normoxic (pimonidazole negative) and hypoxic (pimonidazole positive) regions. The 5hmC signals in each population were compared using ANOVA. PyMT-negative and CD45-positive cells were counted directly. The fraction of pimonidazole and CD31-positive areas was directly quantified using ImageJ across ten images per slide.

Metabolite and protein extraction. For metabolite extractions, 12-well cell culture dishes were placed on ice and washed twice with ice-cold 0.9% NaCl, after which 500 µl of ice-cold 80% methanol was added to each well. Cells were scraped and 500 µl was transferred to a vial on ice. Wells were washed with 500 µl 80% methanol, which was combined with the initial cell extracts. The insoluble fraction was pelleted at 4 °C by a 10-min 21,000g centrifugation. The pellet (containing the proteins) was dried, dissolved in 0.2 N NaOH at 96 °C for 10 min and quantified using a bicinchoninic acid protein assay (BCA, Pierce), whereas the supernatant fraction was processed for metabolite profiling.

Derivation and measurement of metabolites. The supernatant fraction containing the metabolites was transferred to a new vial and dried in a Speedvac. The dried supernatant fraction was dissolved in 45 µl of 2% methoxyamine hydrochloride in pyridine and held for 90 min at 37 °C in a horizontal shaker, followed by derivatization through the addition of 60 µl of N-(tert-butyldimethylsilyl)-n-methyl-trifluoroacetamide with 1% tert-butyldimethylchlorosilane and a 60-min incubation at 60 °C. Samples were subsequently centrifuged for 5 min at 21,000g and 85 µl was transferred to a new vial and analysed using a gas-chromatography based mass spectrometer (triple quadrupole, Agilent) operated in Multiple Reaction Monitoring (MRM) mode.

Analysis of metabolite concentrations. For each sample, metabolite measurements were normalized per sample to the corresponding protein concentration estimates and expressed relative to control-treated samples. Four technical replicates were run for each sample, and the experiment was repeated 4 times using independent samples (*n* = 16). Differences in metabolite concentration were assessed using a two-tailed paired Student's *t*-test or using analysis of variance with post-hoc Tukey HSD when repeated measures were compared.

ROS measurement using 2',7'-dichlorodihydrofluorescein diacetate. MCF7 cells were cultured in 24-well plates and exposed to 21% (control) or 0.5% O₂ (hypoxia) for 24 h. DMEM used for staining was pre-equilibrated to the required O₂ tension, and all steps performed at 21% (control) or 0.5% O₂ (hypoxia) using a glove box. The cells were washed twice with 500 µl DMEM, and incubated for 30 min in 2',7'-dichlorodihydrofluorescein diacetate (DCF-DA; 10 µM) in 500 µl DMEM, keeping 2 wells unstained by DMEM without DCF-DA. Cells were treated with the indicated concentrations of H₂O₂ in DMEM for 30 min at 37 °C, and fixed by adding 33.3 µl of 16% methanol-free paraformaldehyde (Thermo Fisher) for 8 min at room temperature. The fixative was quenched using glycine (150 µM), cells were washed twice in ice-cold PBS, scraped to detach them and transfer them to pre-cooled FACS tubes over cell strainers. Cells were kept on ice until they were analysed by flow cytometry using a FACSVerse (BD Biosciences).

Nuclear ROS measurement using nuclear peroxy emerald 1. MCF7 cells were seeded on 12-well glass-bottom plates and after 24 h exposed to 21% (control) or 0.5% O₂ (hypoxia) for 24 h. PBS used for subsequent staining was pre-equilibrated to the required O₂ tension, and all washing, treatment and staining steps were performed at the appropriate O₂ tension (21% or 0.5%) using a glove box. Cells were loaded with nuclear peroxy emerald 1 (NucPE1; 5 µM)^{38,39} and Hoechst 33342 (10 µg ml⁻¹) in PBS for 15 min at 37 °C. After washing three times in PBS, control cells were incubated with H₂O₂ (0.5 mM in PBS) as a positive control, or with water (control and hypoxia cells) in PBS at 37 °C for 20 min. Cells were washed three times in PBS, placed on ice and immediately imaged by confocal microscopy. The nuclear NucPE1 signal was measured, and averaged across >100 nuclei per replicate using ImageJ. This experiment was repeated 5 times on different days, and signals compared using a *t*-test.

Cell growth measurement using Sulforhodamine B. 5,000 cells/well were seeded in three 96-well plates. After 48 h, one plate was fixed using trichloroacetic acid (3.3% w/v) for 1 h at 4 °C, one plate incubated for 24 h at 37 °C under hypoxic and one under control conditions (0.5% and 21% O₂, respectively). The latter 2 plates were subsequently also fixed using trichloroacetic acid (3.3% wt/vol) for 1 h at

4°C, and all 3 plates were next analysed using the *In vitro* Toxicology Assay Kit, Sulforhodamine B-based (Sigma) as per the manufacturer's instructions. Growth inhibition was calculated as described⁴⁰.

siRNA transfection. siRNA ON-TARGETplus SMART pools (Thermo) were diluted in OptiMEM I reduced serum medium using Lipofectamine RNAiMAX (Life technologies) to reverse-transfect MCF7 cells in 10-cm dishes (for DNA) or 6-well plates (for RNA). Cells were transfected 72 h before RNA and DNA extraction as described.

Hydroxylation assay using nuclear extracts. MCF7 cells were cultured for 24 h under control or hypoxic conditions (21% or 0.5% O₂, respectively), chilled on ice and processed for extraction of nuclear proteins using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific). The activity of control and hypoxic extracts was assessed in parallel using the Colorimetric Epigenase 5mC-Hydroxylase TET Activity/Inhibition Assay Kit (Epigentek) according to manufacturer's instructions. Reactions were allowed to proceed for one hour, after which washing and detection of 5hmC were done according to manufacturer's instructions. Differences between hypoxia and control were analysed using ANOVA, for 5 independent experiments.

DNA hydroxymethylation assay using purified Tet enzyme. The genomic DNA used in this assay was extracted from *Tet* triple-knockout ES cells (G. -L. Xu), and it therefore was devoid of 5hmC⁴¹. To enable efficient denaturation, it was digested using *MseI* before the assay and purified using solid phase reversible immobilisation paramagnetic beads (Agencourt AMPure XP, Beckman Coulter). The assays were performed in Whitley H35 Hypoxystation (don Whitley Scientific) at 37°C, 5% CO₂, N₂, with the following oxygen tensions: 0.1%, 0.3%, 0.5%, 1%, 2.5%, 5%, 10% and 21%. Hypoxystations were calibrated less than 1 month before all experiments. Optimized assay components were as follows: 1.0 µg µl⁻¹ bovine serum albumin (New England Biolabs), 50 mM Tris (pH 7.8), 100 µM dithiothreitol (Life Technologies), 2 ng µl⁻¹ digested gDNA, 250 µM α-ketoglutarate, 830 µM ascorbate, 200 µM FeSO₄ and 45 ng µl⁻¹ Tet1 enzyme (WiseGene). The major assay components (H₂O, BSA and Tris) used for all samples were allowed to pre-equilibrate at 0.1% O₂ for 1 h. These and the remaining assay buffer components (<100 µl) were then pre-equilibrated at the desired oxygen tension for 15 min, and mixed before addition of Tet1 enzyme in a total reaction volume of 25 µl. Reactions were allowed to proceed for 3 min, longer incubations showed a decrease in activity. Reactions were stopped with 80 mM EDTA and stored at -80°C. To measure the resulting 5hmC content of the DNA, reactions were diluted to 100 µl, denatured for 10 min at 98°C and analysed in duplicate using the Global 5-hmC Quantification Kit (Active Motif) following manufacturer's instructions. Michaelis-Menten and Lineweaver-Burk plots and the resulting *K_M* values were estimated using R.

Hypoxia-induced changes in genomic distribution of 5(h)mC in MCF7 cells: DIP-seq. To assess where in the genome the levels of 5mC and 5hmC were altered, we performed DNA immunoprecipitations coupled to high-throughput sequencing (DIP-seq). MCF7 cells were selected for these experiments as they were a cancer cell line with high levels of 5hmC and expression of *TET* genes under control conditions, and a cell growth that is unaffected by hypoxia. This enabled us to study the effects of hypoxia on TET activity in a cell line that shows high endogenous activity, but that is isolated from hypoxia-induced changes in cell proliferation. MCF7 cell culture and DNA extractions were as described for LC-MS analyses. Library preparations and DNA immunoprecipitations were performed as described⁴², using established antibodies targeting 5mC (clone 33D3, Eurogentec,) and 5hmC (Active Motif catalogue number 39791). For 5hmC-DIP-seq, paired barcoded libraries prepared from DNA of hypoxic and control samples were mixed before capture, to enable a direct comparison of 5hmC-DIP-seq signal to the input. A single end of these libraries was sequenced for 50 bases on a HiSeq 2000, mapped using Bowtie and extended for the average insert size (150 bases). Mapping statistics are summarized in Supplementary Information Table 11.

For analysis of sequencing data, MACS peak calling, read depth quantification and annotation with genomic features as annotated in Ensembl build 77 was performed using SeqMonk. Differential (hydroxy-)methylation was quantified by EdgeR⁴³, using either 3 or 5 independent pairs of control and hypoxic samples (for 5hmC-DIP-seq and 5mC-DIP-seq, respectively). These cells were cultured and exposed to hypoxia (0.5% O₂) or control conditions (21% O₂) on different days. Results were reported for 5hmC peak areas that exhibited a change significant at a *P* < 0.05 and 5% FDR.

Target enrichment BS-seq using SeqCapEpi. To confirm enrichment of 5mC at gene promoters using an independent method, DNA libraries were prepared using methylated adapters and the NEBNext DNA library prep master mix set following manufacturer recommendations. Libraries were bisulfite-converted using the Imprint DNA modification kit (Sigma) as recommended, and PCR amplified for 12 cycles using barcoded primers (NEB) and the KAPA HiFi HS Uracil+ ready mix (Sopachem) according to manufacturer's instructions. Fragments were selected from these libraries using the SeqCapEpi CpGiant Enrichment Kit (Roche)

following the manufacturer's instructions, sequenced from both ends for 100 bases on a HiSeq 2000.

For analysing these sequences, sequencing reads were trimmed for adapters using TrimGalore and mapped on a bisulfite-converted human genome (GRCh37) using BisMark. The number of methylated and un-methylated cytosines in captured regions was quantified using Seqmonk for each experiment. Differential methylation of regions of interest was assessed by Fisher's exact test and for 5 independent replicates grown on different days. *t*-scores were averaged following Fisher's method. Mapping statistics are summarized in Supplementary Table 11.

RNA-seq. To assess the effect of the increased 5mC occupancy at gene promoters on their expression, RNA-seq was performed. Briefly, total RNA was extracted using TRIzol (Invitrogen), and remaining DNA contaminants in 17–20 µg of RNA was removed using Turbo DNase (Ambion) according to the manufacturer's instruction. RNA was repurified using RNeasy Mini Kit (Qiagen). Ribosomal RNA present was depleted from 5 µg of total RNA using the RiboMinus Eukaryote System (Life technologies). cDNA synthesis was performed using SuperScript III Reverse Transcriptase kit (Invitrogen). 3 µg of Random Primers (Invitrogen), 8 µl of 5× First-Strand Buffer and 10 µl of RNA mix was incubated at 94°C for 3 min and then at 4°C for 1 min. 2 µl of 10 mM dNTP Mix (Invitrogen), 4 µl of 0.1 M DTT, 2 µl of SUPERase In RNase Inhibitor 20U µl⁻¹ (Ambion), 2 µl of SuperScript III RT (200 U µl⁻¹) and 8 µl of Actinomycin D (1 µg µl⁻¹) were then added and the mix was incubated for 5 min at 25°C, 60 min at 50°C and 15 min at 70°C to heat-inactivate the reaction. The cDNA was purified by using 80 µl (2× volume) of Agencourt AMPure XP and eluted in 50 µl of the following mix: 5 µl of 10× NEBuffer 2, 1.5 µl of 10 mM dNTP mix (10 mM dATP, dCTP, dGTP, dUTP, Sigma), 0.1 µl of RNaseH (10 U µl⁻¹, Ambion), 2.5 µl of DNA Polymerase I Klenov (10 U µl⁻¹, NEB) and the remaining volume of water. The eluted cDNA was incubated for 30 min at 16°C, purified by Agencourt AMPure XP and eluted in 30 µl of dA-Tailing mix (2 µl of Klenow Fragment, 3 µl of 10× NEBNext dA-Tailing Reaction Buffer and 25 µl of water). After 30 min incubation at 37°C, the DNA was purified by Agencourt AMPure XP, eluted in TE buffer and quantified on NanoDrop. Subsequent library preparation was performed using the DNA library prep master mix set and sequencing was performed as described for ChIP-seq. Expression levels (reads per million) of genes displaying significant increases in methylation at their gene promoter, as determined using SeqCapEpi, was compared between control and hypoxic samples using a *t*-test. Mapping statistics are summarized in Supplementary Table 11.

TCGA samples and data analysis. From the TCGA pan-cancer analysis, we selected all solid tumour types for which >100 tumours were available with both gene expression data (RNA-seq) and DNA methylation data (Illumina Infinium HumanMethylation450 BeadChip). These were 408 bladder carcinomas, 691 breast carcinomas, 243 colorectal adenocarcinomas, 520 head and neck squamous cell carcinomas, 290 kidney renal cell carcinomas, 430 lung adenocarcinomas, 371 lung squamous cell carcinomas, and 188 uterine carcinomas, representing in total 3,141 unique patients. Corresponding RNA-seq read counts as well as DNA methylation data from Infinium HumanMethylation450 BeadChip arrays were downloaded from the TCGA server. Breast tumour subtypes were annotated for 208 tumours and, for the remaining tumours, imputed by unsupervised hierarchical clustering of genes in the PAM50 gene expression signature⁴⁴. Other clinical and histological variables were available for >95% of tumours, and missing values were encoded as not available. Gene mutation data was available for 129 bladder carcinomas, 646 breast carcinomas, 200 colorectal adenocarcinomas, 306 head and neck squamous cell carcinomas, 241 kidney renal cell carcinomas, 182 lung adenocarcinomas, 74 lung squamous cell carcinomas, and 3 uterine carcinomas.

Stratification of tumours for hypoxia and proliferation. To identify which of these tumour samples were hypoxic or normoxic, we performed unsupervised hierarchical clustering based a modification (Ward.D of the clusth function in R's stats package) of the Ward error sum of squares hierarchical clustering method⁴⁵, on normalized log₁₀-transformed RNA-seq read counts for 14 genes that make up the hypoxia metagene signature (*ALDOA*, *MIF*, *TUBB6*, *P4HA1*, *SLC2A1*, *PGAM1*, *ENO1*, *LDHA*, *CDKN3*, *TPI1*, *NDRG1*, *VEGFA*, *ACOT7* and *ADM*)²⁵. In each case the top 3 sub-clusters identified were annotated as normoxic, intermediate and hypoxic. To identify which of these tumour samples were high- or low-proliferative, we performed unsupervised hierarchical clustering based a modification (Ward.D of the clusth function in R's stats package) of the Ward error sum of squares hierarchical clustering method⁴⁵, and this for all genes annotated to an established tumour proliferation signature (*MKI67*, *NDC80*, *NUF2*, *PTTG1*, *RRM2*, *BIRC5*, *CCNB1*, *CEP55*, *UBE2C*, *CDC20* and *TYMS*)⁴⁶. Tumours in the top 2 sub-clusters identified were labelled as high- or low-proliferative.

Analysis of the top 1000 CpGs most hypermethylated versus normal tissue. To identify tumour-associated hypermethylation events, we compared 450k methylation data from tumours and normal tissues. All available DNA methylation data from normal tissue (matched or unmatched to tumour samples, on average

59 per tumour type, representing 472 in total, range = 21–160) were downloaded. For each of the 8 tumour types investigated, we selected the top 1,000 CpGs that showed the highest average tumour-associated increases in DNA methylation. Per tumour type, unsupervised hierarchical clustering based on a modification of the Ward error sum of squares hierarchical clustering method (Ward.D of the *clusth* function in R's stats package)⁴⁵ annotated the first 3 clusters identified as having low, intermediate and high hypermethylation. Cluster co-membership for methylation and hypoxia metagene expression were analysed using the Cochran–Armitage test for trend. Analyses using the top 100, 500, 5,000 or 10,000 CpGs yielded near identical results (not shown).

Analysis of hypermethylation events. We next applied a method to identify those CpGs that exhibit exceptional increases in hypermethylation but that are hypermethylated only in a subset of all tumours. Such rare events are typically found in cancer, where hypermethylation inactivates a gene in only a subset of tumours. Hypermethylation of individual CpGs at gene promoters (that is, on average 3.7 CpGs per promoter are represented on the 450K array) in individual tumours was assessed as follows: To achieve a normal distribution, all β -values were transformed to M-values⁴⁷ using $M = \log_2(\beta/(1 - \beta))$. For each tumour type, the mean μ and standard deviation σ of the M value across all control (normoxic) tumours was next calculated, irrespective of mutational status, for each CpG, and used to assign Z-values to each CpG in each tumour using $Z = (M - \mu)/\sigma$. These Z-values describe the deviation in normal methylation variation for that probe. To identify CpGs that display an extreme deviation, we selected those for which the Z-value exceeded 5.6 (that is, $\mu + (5.6 \times \sigma)$), corresponding to a Bonferroni-adjusted *P* value of 0.01; they were considered as hypermethylation events in that particular tumour. This analysis was preferred over Wilcoxon-based models that assess differences in the average methylation level between subgroups, as the latter do not enable the identification or quantification of the rarer hypermethylation events in individual promoters or CpGs.

To identify genes with frequently hypermethylated CpGs in their promoter, the number of hypermethylation events in that promoter was counted in all tumours, and contrasted to the expected number of hypermethylation events in that promoter (that is, the general hypermethylation frequency multiplied by the number of CpGs assessed in that promoter multiplied by the number of tumours) using Fisher's exact test. Genes with an associated Bonferroni-adjusted *P* value below 0.01 were retained and considered as frequently hypermethylated in that tumour type.

To assess what fraction of these hypermethylation events are hypoxia-related, we assumed that the fraction of events detected under normoxia was hypoxia-unrelated, and that all excess events detected in intermediate and hypoxic tumours were hypoxia-related. For example, in 691 breast carcinomas, 0.25% of CpGs were hypermethylated in 251 normoxic tumours, 0.81% in 350 intermediate and 1.40% in 90 hypoxic tumours. So, 0.56% and 1.15% of hypermethylation events in respectively intermediate and hypoxic tumours were hypoxia-related. Taking into account the number of tumours, 0.25% of hypermethylation events (that is, $(0.25\% \times 251 + 0.25\% \times 350 + 0.25\% \times 90)/691$) are not hypoxia-related, and 0.43% are hypoxia related (that is, $(0\% \times 251 + 0.56\% \times 350 + 1.15\% \times 90)/691$). So, 63% of all hypermethylation events combined (that is, $0.43/(0.43 + 0.25)$) are hypoxia related. To assess the contribution of hypoxia to hypermethylation relative to other covariates, partial R^2 values were calculated for the contribution of each covariate in a linear model, and compared to the total R^2 achieved by the model.

To identify genes with CpGs in their promoter that are more frequently hypermethylated in hypoxic than normoxic tumours, the number of hypermethylation events in that promoter was counted in all hypoxic tumours, and contrasted to the number found in normoxic tumours. Differences in frequencies were assessed using Fisher's exact test, and genes with a Bonferroni-adjusted *P* < 0.01 were retained and considered hypermethylated upon hypoxia. Enrichment of ontologies associated with these genes was assessed using Fisher's exact test as implemented in R's topGO package.

Analysis of the effect of hypermethylation events on the expression of associated genes. To enable a direct comparison between the expression of different genes, we transformed gene expression values (reads per million) to their respective z-scores. To assess the impact of hypermethylation events on the expression of associated genes, we compared the expression z-scores of all frequently hypermethylated genes that contain one or more hypermethylation events in their promoter (on average each promoter contains 3.7 CpGs; if one of these is hypermethylated the associated gene is considered hypermethylated in that particular tumour), to the expression of all frequently hypermethylated genes that do not contain a hypermethylation event. The effect of hypermethylation on gene expression was plotted for the 8 main tumour types stratified into normoxic, intermediately hypoxic and hypoxic tumours, and for glioblastomas was stratified into normoxic, intermediately hypoxic, hypoxic and *IDH*-mutant tumours (*n* = 4). The difference in expression z-scores between genes not carrying and carrying a hypermethylation event in their promoter was assessed using a *t*-test.

Analysis of the effect of frequent mutations on the occurrence of hypermethylation events. To assess the impact of somatic mutations on hypoxia-associated hypermethylation frequencies, we analysed the top 20 genes described to be most frequently mutated in the pan-cancer analysis²⁴, and supplemented this list with genes known to cause hypermethylation upon mutation (that is, *IDH1*, *IDH2*, *SDHA*, *FH*, *TET1*, *TET2* and *TET3*). Mutations in *IDH1* and *IDH2* were retained if they respectively affected amino acid R132, and amino acids R140 or R172. Mutations in other genes were scored using Polyphen, and only mutations classified as probably damaging were retained. 7 mutations were found in lung tumours, 3 mutations in colorectal tumours, 8 mutations in breast tumours and 6 mutations (all *IDH1*^{R132}) in glioblastomas. None of these mutations were enriched in hypoxic subsets. In multivariate analyses of variance, in each of the tumour types analysed, mutations in these genes were significantly associated with increased hypermethylation frequencies. Hypoxia was independently and significantly associated with the hypermethylation frequency.

Correlation between hypermethylation and expression of TET or DNMT enzymes. Gene expression values (reads per million) of DNMT and TET enzymes were determined for each tumour using available RNA-seq data. The number of hypermethylation events at significantly hypermethylated genes in each tumour was determined as described above. Hypermethylation in each tumour was subsequently correlated to *TET* or *DNMT* gene expression in that tumour, correcting for hypoxia and proliferation status using ANOVA.

5mC and 5hmC profiling using 450k arrays for 24 lung tumours. Newly diagnosed and untreated non-small-cell lung cancer patients scheduled for curative-intent surgery were prospectively recruited. Included subjects had a smoking history of at least 15 pack-years. The study protocol was approved by the Ethics Committee of the University Hospital Gasthuisberg (Leuven, Belgium). All participants provided written informed consent. In the framework of a different project⁴⁸, RNA-seq was performed on 39 tumours from these patients. Gene expression for these samples was clustered for their hypoxia metagene signature²⁵. This yielded 2 clear clusters, containing 24 and 15 normoxic and hypoxic tumours, respectively. Twelve samples were randomly selected from each cluster, in order to perform 5hmC and 5mC profiling.

Illumina Infinium HumanMethylation450 BeadChips. For TAB–ChIP, DNA was glycosylated and oxidized as described⁴⁹, using the 5hmC TAB–Seq Kit (WiseGene). Subsequently, bisulfite conversion, DNA amplification and array hybridization were done following manufacturer's instructions.

Analysis of TAB–ChIP and BS–ChIP. Data processing was largely as described⁵⁰. In brief, intensity data files were read directly into R. Each sample was normalized using Subset–quantile within array normalization (SWAN) for Illumina Infinium HumanMethylation450 BeadChips⁴⁹. Batch effects between chips and experiments were corrected using the runComBat function from the ChAMP bioconductor package⁵¹. For obtaining 5mC-specific beta values, TAB–ChIP generated normalized beta values were subtracted from the standard 450K generated normalized beta values, exactly as described⁵⁰.

Mouse cancer models. All the experimental procedures were approved by the Institutional Animal Care and Research Advisory Committee of the KU Leuven.

Hypoxia induction using sFlk1-overexpression. For sFlk1-overexpression studies, male Tg(MMTV–PyMT) FVB mice were intercrossed with wild-type FVB female mice. Female pups of the Tg(MMTV–PyMT) genotype were retained, and tumours allowed to develop for 9 weeks. Subsequently, 2.5 µg of plasmid (sFlk1-overexpressing or empty vector; randomly assigned within litter mates) per gram of mouse body weight was introduced in the bloodstream using hydrodynamic tail vein injections⁵². sFlk1 overexpression was monitored at 4 days after injection and at the day of killing (18 days after the injection), by eye bleeds followed by an ELISA assay for sFlk1 (R&D Systems) in blood plasma. At 12 weeks of age, mice were killed and mammary tumours collected were blinded for treatment.

Hypoxia reduction using Phd2 haploinsufficiency. For the *Phd2*^{+/-} experiments, male Tg(MMTV–PyMT) FVB mice were intercrossed with female *Phd2*^{-/+} mice, yielding litters of which half have either a Tg(MMTV–PyMT) genotype or a Tg(MMTV–PyMT);*Phd2*^{-/+} genotype. For the *Phd2*^{WT/Δ} experiments, male Tg(MMTV–PyMT) FVB mice were intercrossed with female *Tie2*–Cre;*Phd2*^{WT/Δ} mice as described²⁷, yielding litters of which half have either a Tg(MMTV–PyMT);*Phd2*^{WT/WT} genotype or a *Tie2*–cre;Tg(MMTV–PyMT);*Phd2*^{-/+} genotype. At 16 weeks of age, female mice were killed and mammary tumours collected.

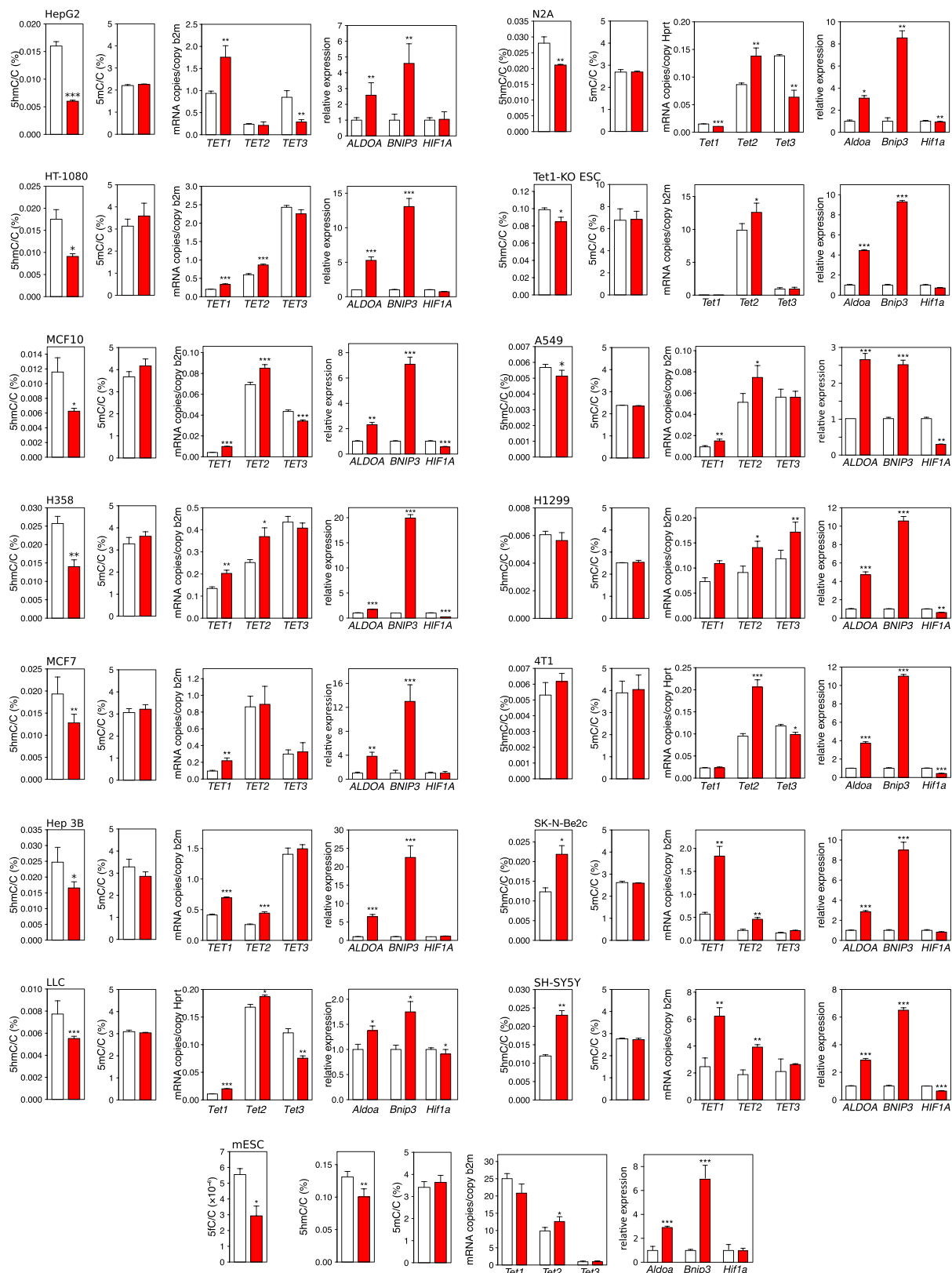
qPCR analysis of expression of Tets and marker genes. RNA was extracted from fresh-frozen tumours (stored at –8 °C) using TRIzol (Life Technologies), and converted to cDNA and quantified as described for the cell lines. TaqMan probes and primers (IDT or Life Technologies) are described under Supplementary Table 12. **TAB–seq of PyMT tumours.** TAB–seq libraries were prepared as described, using the 5hmC TAB–Seq Kit (WiseGene). DNA was bisulfite-converted using the EZ DNA Methylation–Lightning Kit (Zymo Research) and sequenced as described for SeqCapEpi experiments. Reads were mapped to the mouse genome

(build Mm9) and further data processing was as for SeqCapEpi experiments. DNA from 3 independent tumours was selected per condition. TET oxidation efficiency was required to exceed 99.5% as estimated using a fully CG-methylated plasmid spike-in, 5hmC protection by glycosylation was 65% as estimated using a fully hydroxymethylated plasmid spike-in, bisulfite conversion efficiencies were estimated to exceed 99.8% based on nonCG methylation (equal to percentage hypermethylated CpG). Mapping statistics are summarized in Supplementary Table 11.

Targeted deep BS-seq. As no mouse capture kit was available for targeted BS-seq, a specific ampliconBS was developed for a set of 15 tumour suppressor gene promoters and 5 oncogene promoters. More specifically, DNA was bisulfite-converted using the Imprint DNA modification kit and amplified using the MegaMix Gold 2× mastermix and validated primer pairs. Per sample, PCR products were mixed to equimolar concentrations, converted into sequencing libraries using the NEBNext DNA library prep master mix set and sequenced to a depth of ~500×. Mapping and quantification were done as described for SeqCapEpi. The average and variance of methylation level M values in normal mammary glands were used as baseline, and amplicons displaying over 3 standard deviations more methylation (FDR-adjusted $P < 0.05$) than this baseline were called as hypermethylated. At least 9 different tumours, each from different animals, were profiled per genotype or treatment, and differences in hypermethylation frequencies between sets of tumours were assessed using Mann–Whitney’s U-test.

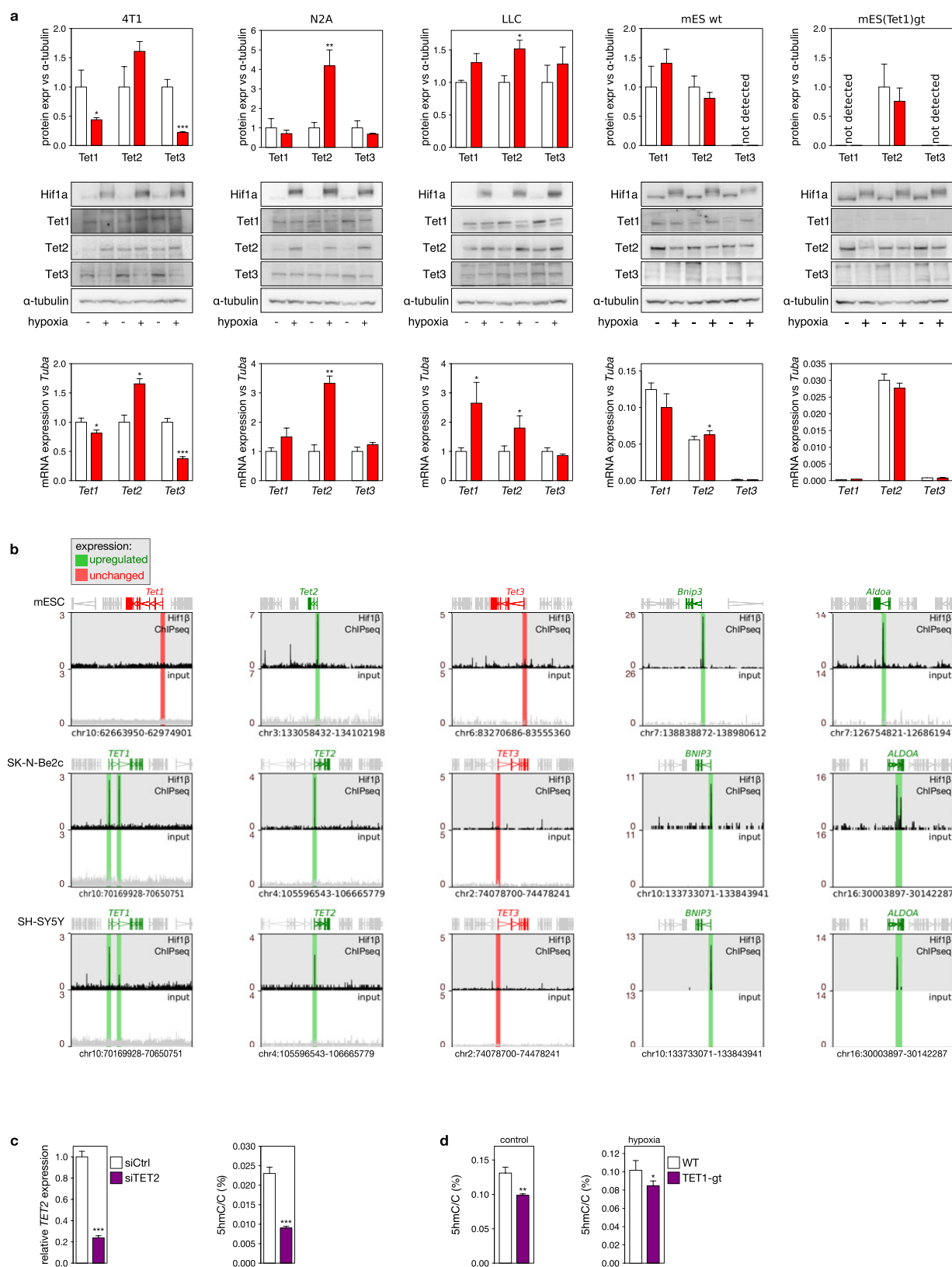
Statistics. Data entry and analysis were performed in a blinded fashion. Statistical significance was calculated by two-tailed unpaired *t*-test (Excel) or analysis of variance (R) when repeated measures were compared. Data were tested for normality using the D’Agostino–Pearson omnibus test (for $n > 8$) or the Kolmogorov–Smirnov test (for $n \leq 8$) and variation within each experimental group was assessed. Data are presented as mean \pm s.e.m. DNA methylation and RNA-seq gene expression data distributions were transformed to a normal distribution by conversion to M values and \log_2 transformation respectively. Sample sizes were chosen based on prior experience for *in vitro* and mouse experiments, or on sample and data availability for human tumour analyses. Other statistical methods (mostly related to specific sequencing experiments) are described together with the experimental details in other sections of the methods.

33. Sermeus, A. *et al.* Hypoxia induces protection against etoposide-induced apoptosis: molecular profiling of changes in gene expression and transcription factor activity. *Mol. Cancer* **7**, 27 (2008).
34. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protocols* **3**, 1101–1108 (2008).
35. Majmundar, A. J., Wong, W. J. & Simon, M. C. Hypoxia-inducible factors and the response to hypoxic stress. *Mol. Cell* **40**, 294–309 (2010).
36. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. *Curr. Protoc. Bioinformatics* **Chapter 2**, Unit 2.14, (2011).
37. Durand, R. E. & Raleigh, J. A. Identification of nonproliferating but viable hypoxic tumor cells *in vivo*. *Cancer Res.* **58**, 3547–3550 (1998).
38. Lippert, A. R., Van de Bittner, G. C. & Chang, C. J. Boronate oxidation as a bioorthogonal reaction approach for studying the chemistry of hydrogen peroxide in living systems. *Acc. Chem. Res.* **44**, 793–804 (2011).
39. Dickinson, B. C., Tang, Y., Chang, Z. & Chang, C. J. A nuclear-localized fluorescent hydrogen peroxide probe for monitoring sirtuin-mediated oxidative stress responses *in vivo*. *Chem. Biol.* **18**, 943–948 (2011).
40. Vichai, V. & Kirtikara, K. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat. Protocols* **1**, 1112–1116 (2006).
41. He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
42. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protocols* **7**, 617–636 (2012).
43. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
44. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
45. Murtagh, F. & Legendre, P. Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion? *J. Classif.* **31**, 274–295 (2014).
46. Nielsen, T. O. *et al.* A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **16**, 5222–5232 (2010).
47. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
48. Wauters, E. *et al.* DNA methylation profiling of non-small cell lung cancer reveals a COPD-driven immune-related signature. *Thorax* **70**, 1113–1122 (2015).
49. Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protocols* **7**, 2159–2170 (2012).
50. Nazor, K. L. *et al.* Application of a low cost array-based technique - TAB-Array - for quantifying and mapping both 5mC and 5hmC at single base resolution in human pluripotent stem cells. *Genomics* **104**, 358–367 (2014).
51. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430 (2014).
52. Liu, F., Song, Y. & Liu, D. Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. *Gene Ther.* **6**, 1258–1266 (1999).
53. Pelicano, H., Carney, D. & Huang, P. ROS stress in cancer cells and therapeutic implications. *Drug Resist. Updat.* **7**, 97–110 (2004).
54. Booth, M. J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).



Extended Data Figure 1 | Hypoxia-induced changes in 5hmC, 5mC and *TET* expression. Global 5hmC/C and 5mC/C content of DNA, *TET1*, *TET2* and *TET3* mRNA expression and hypoxia marker gene expression in 15 cell lines grown for 24 h under normoxic (21% O₂, white) or hypoxic (0.5% O₂, red) conditions. *TET* mRNA copy number is expressed relative to *B2M* for human cell lines (HepG2, HT-1080, MCF10A, H358, MCF7, Hep3B, A549, H1299, SK-N-Be2c and SHSY5Y), and to *Hprt* for mouse cell lines (LLC, N2A, 4T1, mES and Tet1-KO ES cells). Shown are cell lines derived from liver cancer (HepG2 and Hep3B), lung cancer (H358, A549, H1299 and LLC),

breast cancer (MCF7 and 4T1), fibrosarcoma (HT1080), neuroblastoma (SK-N-Be2c and SHSY5Y), normal breast epithelium (MCF10A) and the inner cell mass of blastocyst-stage mouse embryos (mES and Tet1-KO ES cells). *ALDOA* and *BNIP3* are expected to be increased, and *HIF1A* to be decreased upon hypoxia. The global 5fC content of ES cells is depicted, but was undetectable in cancer cell lines. Bars represent the mean \pm s.e.m. of five different replicate samples. DNA and RNA from these replicates were extracted from cells derived from the same stock vial but grown on different days. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by paired *t*-tests.

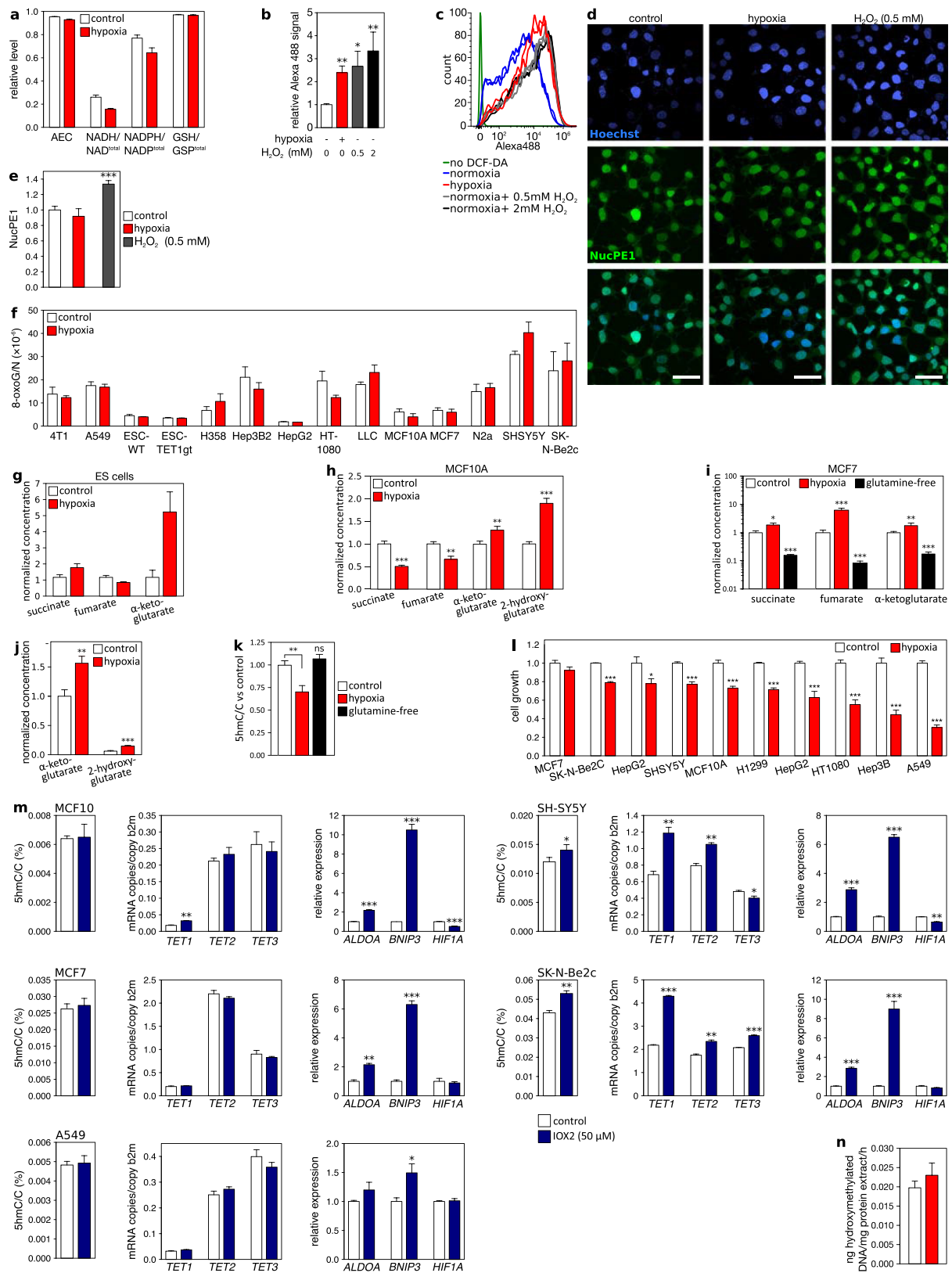


Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Impact of hypoxia on TET expression.

a, Changes in Tet1, Tet2 and Tet3 expression in mouse cell lines, at the protein level (top row, $n = 6$) and the mRNA level (bottom row, $n = 5$). Middle row: representative immunoblot images of Hif1a, Tet1, Tet2 and Tet3. α -Tubulin serves as loading control, and expression of the corresponding coding gene (*Tuba1a*) was used to normalize mRNA expression, enabling a direct comparison of relative protein and relative mRNA expression changes. For the same reason, mRNA expression was depicted relative to control conditions, in contrast to the absolute levels shown in Extended Data Fig. 1. Changes in Tet mRNA and protein expression correlate strongly (Pearson's $R = 0.855$, $P = 4 \times 10^{-4}$). For example, both 4T1 and N2A cells displayed increased Tet2 expression at the protein and mRNA level. Likewise, ES cells showed no pronounced changes at the protein or mRNA level. The overall expression of Tet enzymes was not altered in any of these cell lines. For gel source data,

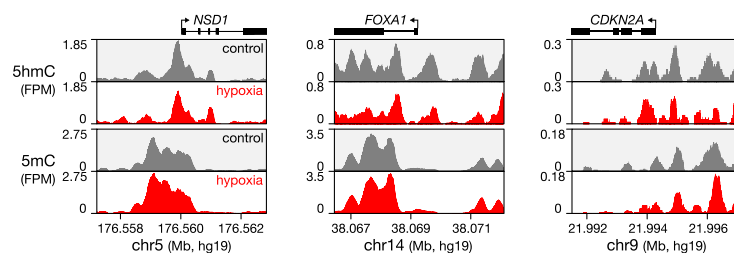
see Supplementary Fig. 1. **b**, Hif1 β ChIP-seq at the promoters of *TET1*, *TET2* and *TET3* and at hypoxia markers genes (*Bnip3* and *Aldoa*), with peaks or promoter regions highlighted using coloured boxes. Green and red boxes correspond to overexpression and no overexpression (specified in the figure panel) of the corresponding gene, respectively, as determined using TaqMan in Extended Data Fig. 1. Scale: reads per million reads and per base pair. **c**, Left, *TET2* expression in MCF7 cells transfected with control (white) or *TET2*-targeting (purple) siRNAs. Right, corresponding 5hmC levels as determined using LC-MS. **d**, 5hmC levels as determined by LC-MS, in wild-type (white) and *Tet1*-knockout (purple) ES cells grown under 21% (left) and 0.5% (right) O₂ tensions. Bars in **c** and **d** represent the mean \pm s.e.m. of five replicate samples from cells derived from the same stock vial but grown on different days. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by paired t -tests (**a**, **c**, **d**).



Extended Data Figure 3 | See next page for caption.

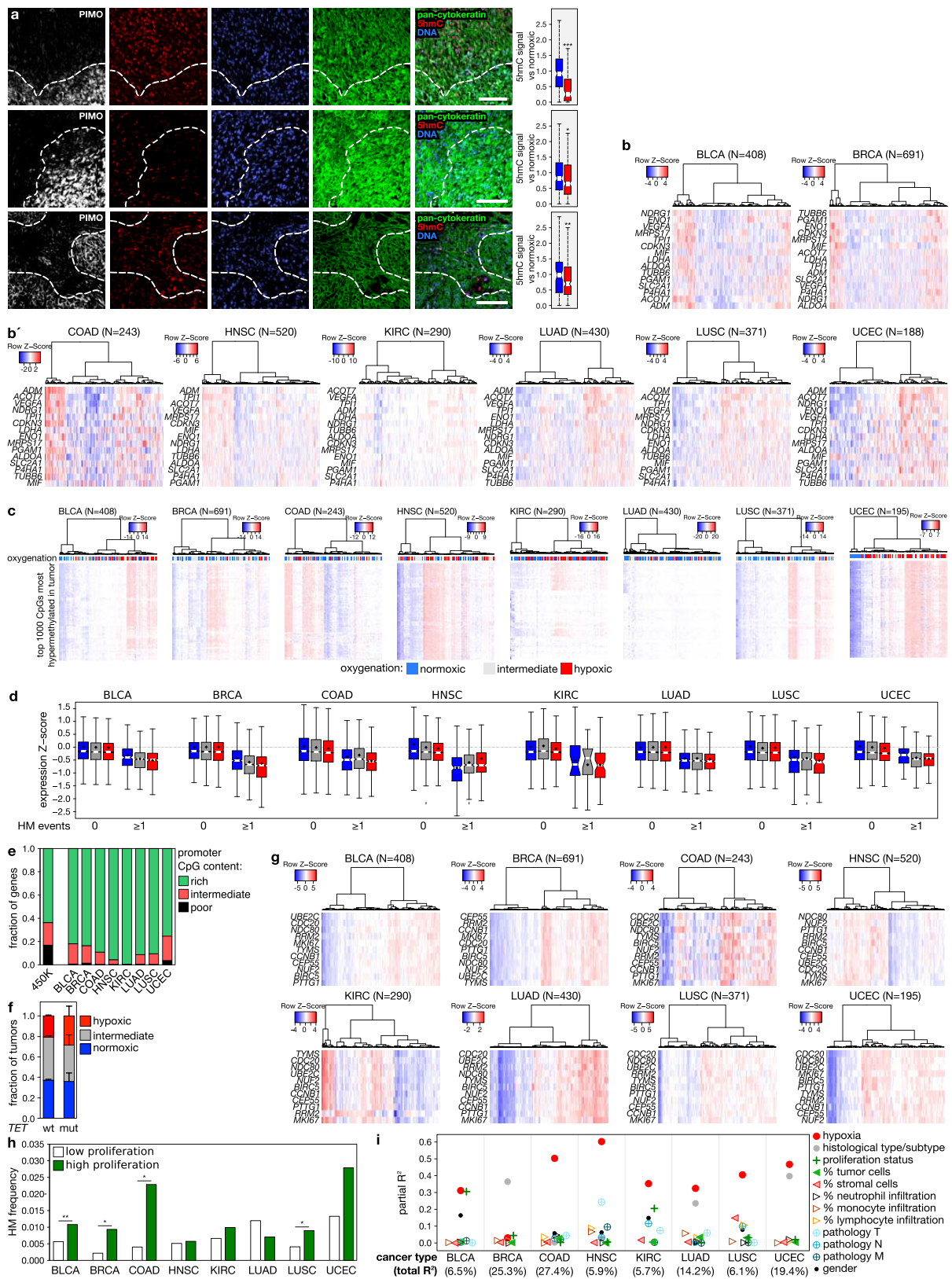
Extended Data Figure 3 | Effects secondary to hypoxia. a–e, ROS production and redox state of MCF7 cells cultured for 24 h under control (21% O₂, white) or hypoxic (0.5% O₂, red) conditions. Shown are capillary gas chromatography mass spectrometry (GC–MS) quantifications of changes in the cellular energy state as represented by the adenylate energy charge (AEC) (calculated as $[\text{ATP} + 0.5 \times \text{ADP}]/[\text{ATP} + \text{ADP} + \text{AMP}]$) (**a**); the reducing equivalents of the cell as represented by the relative NADH and NADPH levels (calculated as $\text{NADH}/[\text{NAD}^+ + \text{NADH}]$ and $\text{NADPH}/[\text{NADP}^+ + \text{NADPH}]$); and the reductive capacity of the cell as represented by the levels of glutathione (calculated as $\text{GSH}/[\text{GSH} + \text{GSSG} \times 2]$). **b, c,** Quantification (**b**) and representative FACS intensity traces (**c**) of total ROS levels in MCF7 cells exposed to hypoxia or H₂O₂, as assessed using 2',7'-dichlorodihydrofluorescein diacetate (DCF-DA). **d,** Nuclear ROS in MCF7 cells as assessed using the nuclear peroxyl emerald 1 probe (NucPE1)³⁹. MCF7 cells were exposed to 21% (control) or 0.5% (hypoxia) O₂ for 24 h, after which live cells were loaded with NucPE1 (5 μM) and Hoechst 33342 (10 μg ml⁻¹) in O₂ pre-equilibrated PBS for 15 min. After washing, control cells were incubated with H₂O₂ (0.5 mM in PBS) as a positive control, or with water (control and hypoxia cells) in PBS for 20 min. Cells were washed again and immediately imaged by confocal microscopy. Representative images are shown. Scale bar, 50 μm. **e,** The nuclear NucPE1 signal, averaged across >100 nuclei and expressed relative

to control conditions. **f,** LC–MS quantification of 8-oxoG concentrations in DNA of cells lines cultured for 24 h under control (21% O₂, white) and hypoxic (0.5% O₂, red) conditions. 8-oxoG serves as a marker of nuclear ROS⁵³. **g–i,** GC–MS quantification of changes in the indicated metabolite levels in mouse ES cells (**g**), MCF10A cells (**h**) and MCF7 cells (**i**) grown for 24 h under control (21% O₂, white), hypoxic (0.5% O₂, red) or glutamine-free (21% O₂, black) conditions. **j,** Quantities of α-ketoglutarate and 2-hydroxyglutarate in MCF7 cells, expressed relative to α-ketoglutarate levels in MCF7 cells grown under control conditions (21% O₂). **k,** LC–MS quantification of 5hmC levels in response to hypoxia (0.5% O₂) and glutamine-free culture conditions. **l,** Growth of cell lines cultured for 24 h under control (21% O₂, white) and hypoxic (0.5% O₂, red) conditions, as assessed using a sulforhodamine B colorimetric assay. Changes in cell density after 24 h are depicted relative to control conditions (21% O₂). **m,** IOX2-induced changes in the global 5hmC content of DNA, in *TET* mRNA expression and in hypoxia marker gene expression of five cell lines treated for 24 h with DMSO (carrier, white) or IOX2 (50 μM, blue). **n,** 5mC hydroxylation activity of nuclear lysates from MCF7 cells grown for 24 h under 21% or 0.5% O₂ (white or red). Bars represent the mean ± s.e.m. of 5 (**b, k, m**), 6 (**a, c–e**), 16 (**g–j**) or 24 (**l**) samples prepared on different days. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 by *t*-test (**b, e, h–m**).



Extended Data Figure 4 | Genomic profiles of 5mC and 5hmC. Shown are results from DIP-seq of DNA from MCF7 cells cultured for 24 h under 21% or 0.5% O₂ (control and hypoxia), with examples of 5hmC-DIP-seq (top) and 5mC-DIP-seq (bottom) read depths (FPM, fragments per base pair per million fragments) at regions surrounding the transcription start site of *NSD1*, *FOXA1* and *CDKN2A*. These show 5hmC loss (FDR < 5%)

and a 5mC gain that is more subtle, perhaps because the resolution of 5mC-DIP-seq is limiting: regions rich in 5hmC tend to be poorer in 5mC⁵⁴, and thus have less substrate available for pull-down. 5mC-DIP-seq moreover captures all methylated sites, so most of the 5mC-DIP-seq signal does not derive from sites that are actively turning over 5hmC.

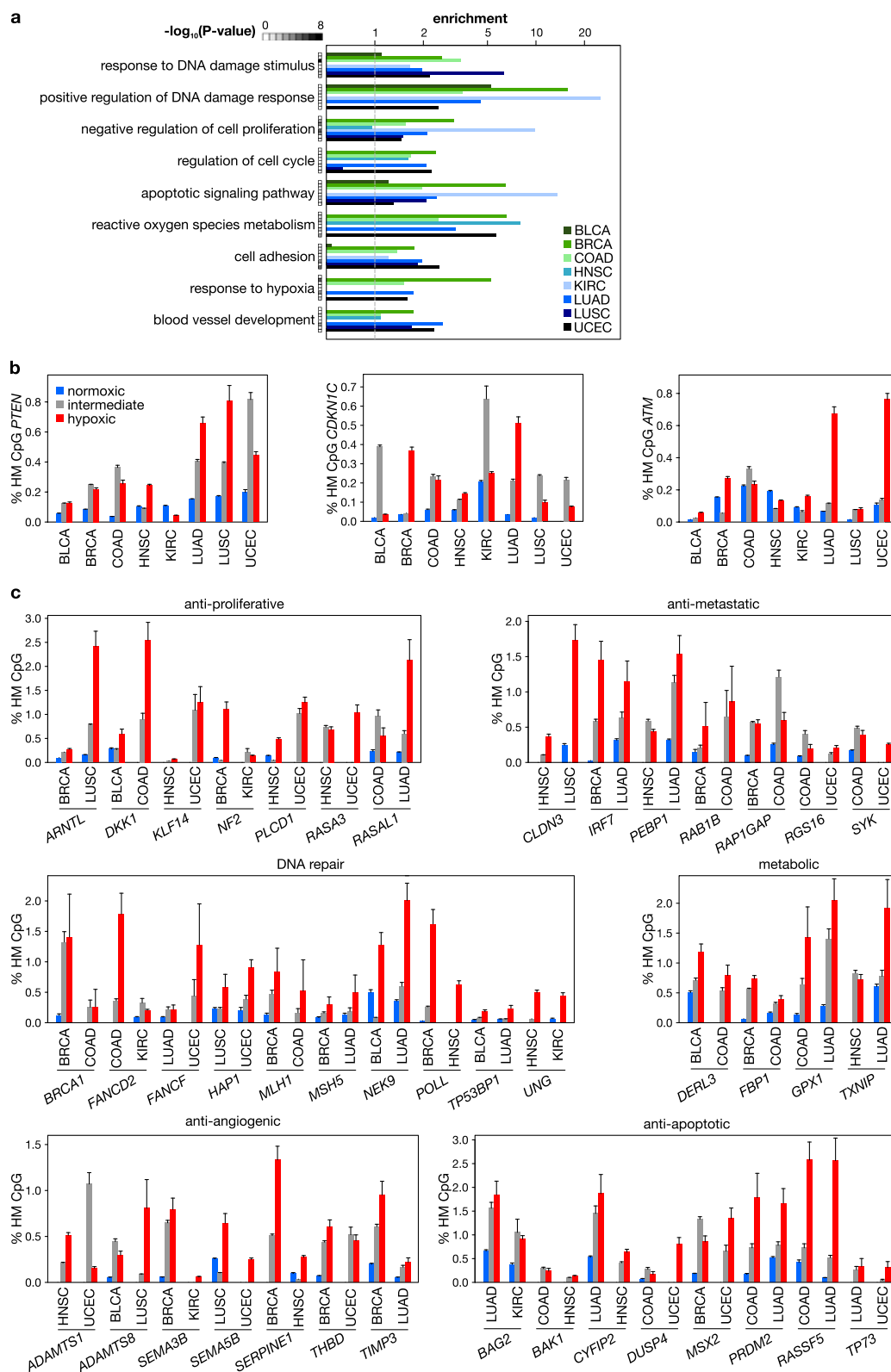


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Effect of hypoxia on hypermethylation frequency in tumours.

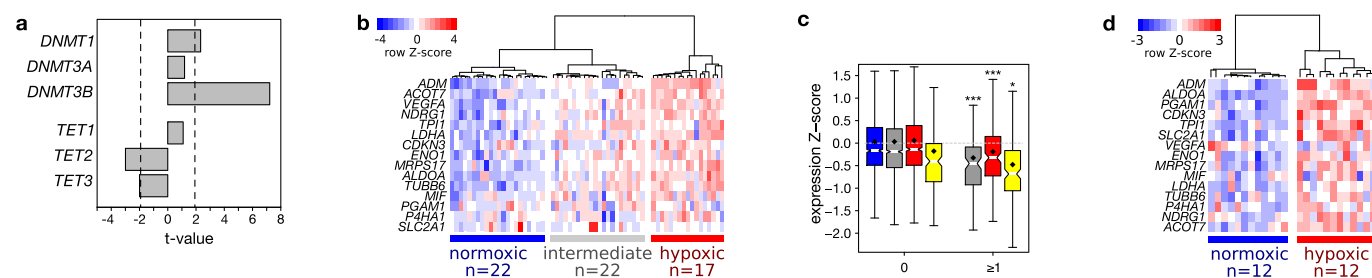
a, Immunofluorescence analysis of patient-derived tumour xenografts, stained for pimonidazole (white), 5hmC (red), DNA (propidium iodide, blue) and pan-cytokeratin (green). Shown are representative images of a breast and two endometrial tumour xenografts. The inset on the right shows box plots illustrating the signal in normoxic pimonidazole-negative nuclei (blue), and in hypoxic pimonidazole-positive nuclei (red). **b**, Hypoxia marker gene expression clusters, with the first three clusters used to define normoxic, intermediate and hypoxic tumours. **c**, Unsupervised clustering of 1,000 CpGs showing the highest average methylation increase in tumour versus corresponding normal tissues. The first three clusters were used to define tumours of low, intermediate and high hypermethylation. The colour bar above the clusters annotates each tumour as normoxic, intermediate or hypoxic, as determined in **b**. **d**, Box plots showing the relative expression (z-score) of genes in tumours in which they have either 0 or ≥ 1 hypermethylation event in their promoter, stratified into normoxic, intermediate and hypoxic tumours (blue, grey and red, respectively). Diamonds indicate mean, box plot wedges indicate $2\times$ the standard error of the median. Genes with ≥ 1 hypermethylation events in their promoters have a lower

average expression level ($P < 0.01$ for each tumour type). **e**, Fraction of genes having a promoter that is rich, intermediate or poor in CpGs, out of all gene promoters that are assessed on the 450k array, and out of all gene promoters that are frequently hypermethylated in the indicated tumour types. **f**, Fraction of 1,742 *TET* wild-type tumours and 39 *TET* mutant that are normoxic, intermediate and hypoxic. $P > 0.2$ for all comparisons. **g**, Cell proliferation marker gene⁴⁶ expression clusters, with the first two clusters used to define high-proliferative and low-proliferative tumours. **h**, hypermethylation frequencies in low- and high-proliferative tumours, with asterisks representing P values from linear models correcting for variables specified in Supplementary Table 8. **i**, Partial correlation coefficient (partial R^2) estimates of the relative contribution of tumour characteristics (annotated in TCGA) to the variance in hypermethylation observed in these tumours. Partial R^2 values were obtained from linear model estimation using ordinary least squares, and expressed as a fraction of the total variance (that is, total R^2) explained by the model when taking into account all indicated variables, as indicated between brackets under each tumour type. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by t -test (**a**) or by generalized linear model (**h**).



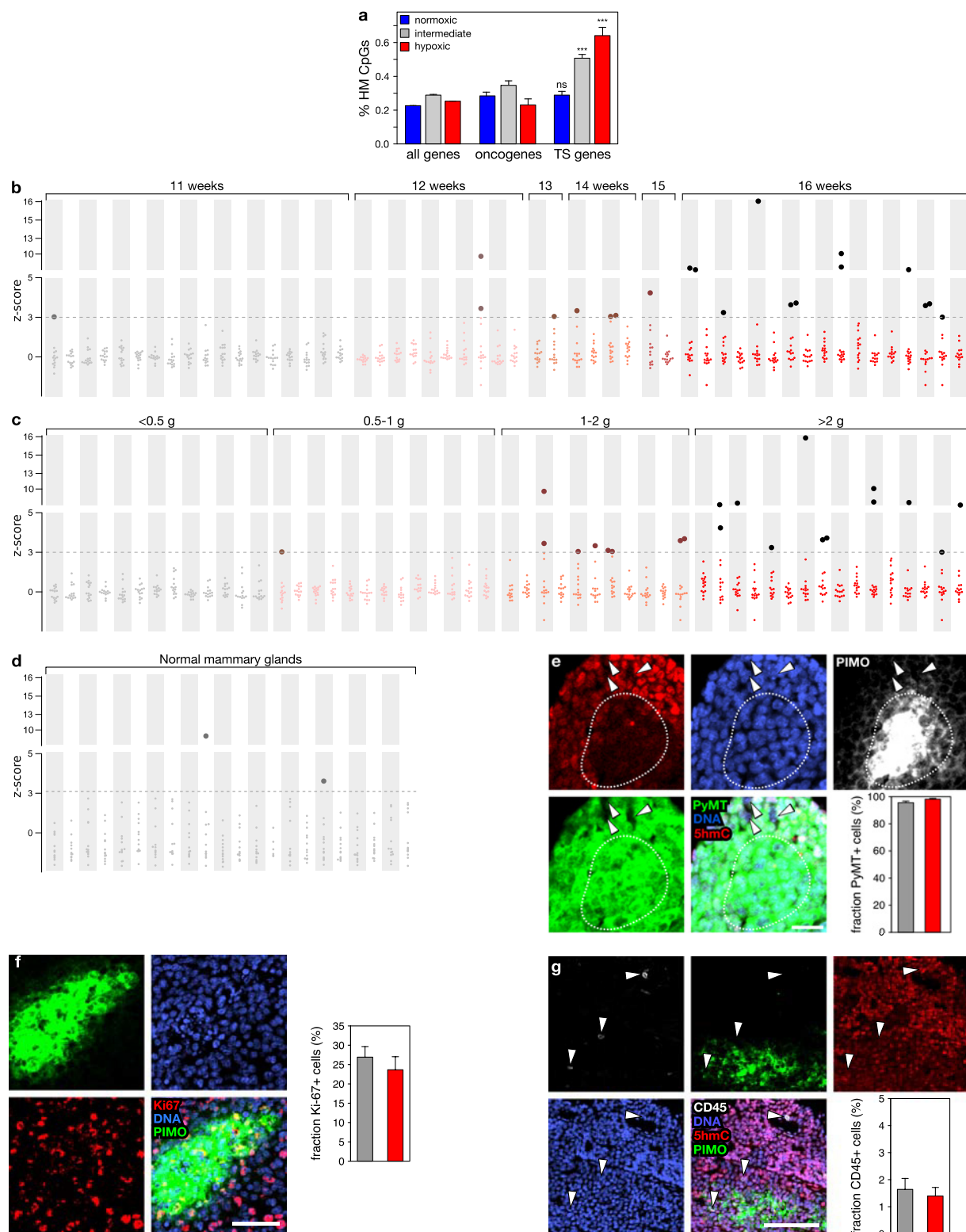
Extended Data Figure 6 | Functional annotation of genes more frequently hypermethylated in hypoxic tumours. a, Ontology terms enrichment analysis of genes that are more frequently hypermethylated at their gene promoters in hypoxic than normoxic tumours, for eight tumour types characterized in the TCGA pan-cancer effort. A representative set of terms is displayed, selected from terms enriched in most tumour types. *P* values as defined by the grey-scale insert. Enrichment calculated

using topGO. **b,** Selected examples of hypermethylation frequencies in the promoter of key TSGs (*PTEN*, *CDKN1C*, *ATM*) more frequently hypermethylated in normoxic than hypoxic tumours. **c,** Hypermethylation frequency in the promoter of selected genes involved in the processes indicated. *P* < 0.05 for all genes (asterisks are not displayed). Bars in **b** and **c** represent the hypermethylation frequency \pm s.e.m. *P* values in **a** by Fisher's exact test.



Extended Data Figure 7 | Effect of hypoxia on TET activity in human tumours. **a**, The t -value of correlation between hypermethylation and expression of *TET* or *DNMT* genes across 3,141 tumours of 8 tumour types (bladder, breast, colorectal, head and neck, kidney, lung adenocarcinoma, lung squamous, and uterine carcinoma) profiled in TCGA for gene expression and DNA methylation, while correcting for tumour type, hypoxia and proliferation. The dotted line represents $P < 0.05$, negative t -values represent inverse correlations. **b**, Hypoxia metagene signature applied to 63 glioblastoma multiforme tumours from TCGA. **c**, Boxplots showing the relative expression (Z-score) of genes in

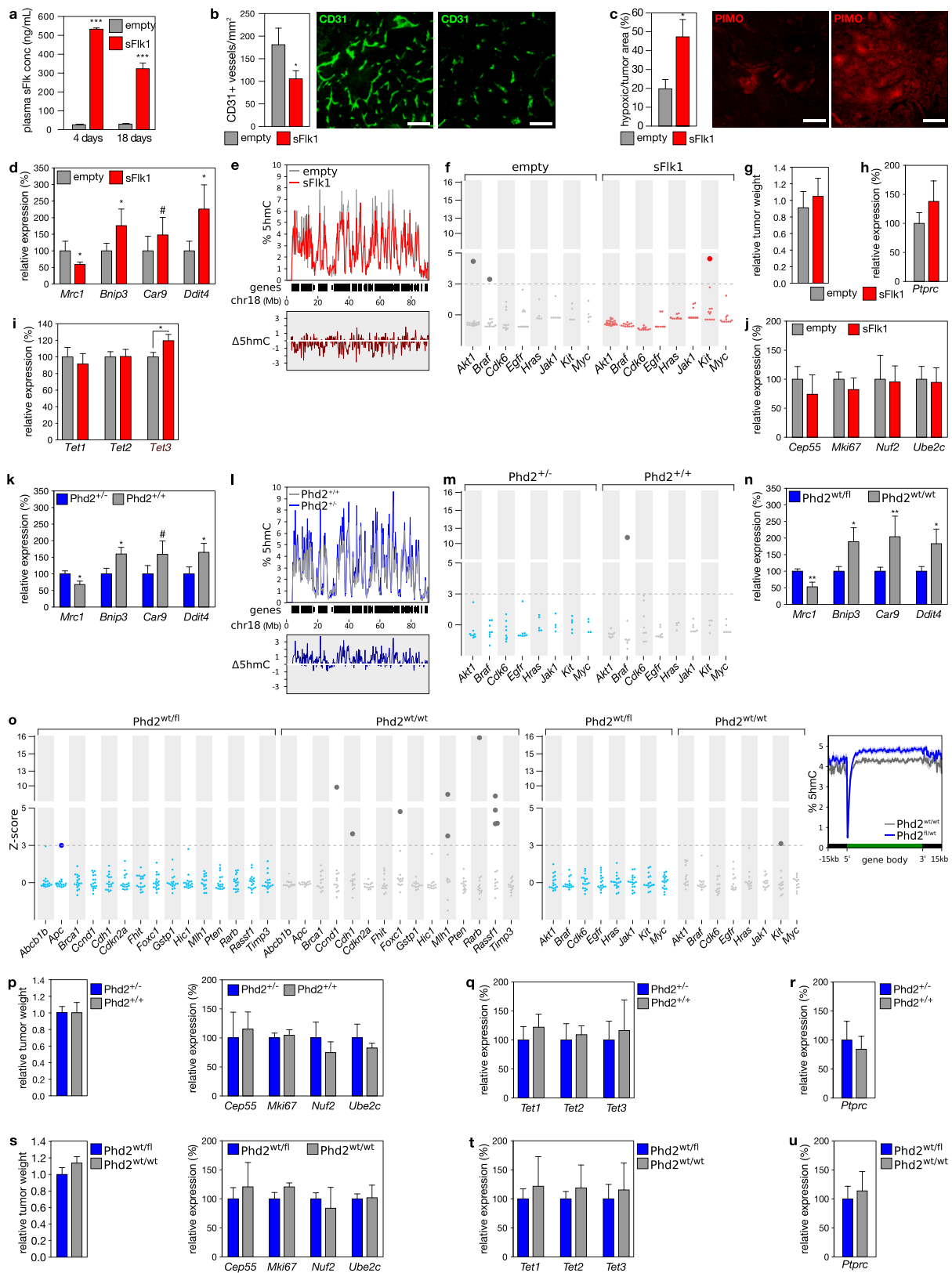
tumours in which they have either 0 or ≥ 1 hypermethylation event in their promoter, stratified into wild-type *IHD1* tumours that are normoxic ($n = 19$), intermediate ($n = 21$) and hypoxic ($n = 17$) (blue, grey and red, respectively), and *IDH1*^{R138}-mutated tumours ($n = 4$, yellow). Diamonds indicate mean, box plot wedges indicate $2 \times$ the standard error of the median. Genes with ≥ 1 hypermethylation events in their promoters have a lower average expression level. No hypermethylation events were detected in wild-type *IHD1* normoxic tumours. **d**, Hypoxia metagene signature applied to 12 normoxic and 12 hypoxic non-small-cell lung tumours. * $P < 0.05$, *** $P < 0.001$ by t -test (**c**).



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | 5hmC, hypoxia and TSG hypermethylation in mouse breast tumours. **a**, Frequency of hypermethylation events in the promoters of all genes, all oncogenes and all TSGs as annotated²⁸, in 695 human breast tumours available through TCGA and stratified into normoxic, intermediate and hypoxic subsets. **b, c**, DNA was extracted from 53 tumours developing in MMTV-PyMT mice of the indicated ages (**c**) or weights (**d**) and sequenced to a depth of $\sim 500\times$. Plotted are z-scores of hypermethylation (y axis, exponential) for 15 TSGs, relative to the tumours from 11-week-old mice. The dotted line represents the threshold for a Bonferroni-adjusted $P < 0.05$, and bold darker dots are used for tumours displaying significantly increased hypermethylation events. **d**, DNA extracted from 20 normal mammary glands from 14-week-old mice, PCR-amplified for 15 TSGs and sequenced to a depth of $\sim 500\times$. Plotted are z-scores of hypermethylation relative to 11-week-old tumours. **e**, Staining of PyMT tumours for 5hmC (red), DNA (propidium iodide, blue), pimonidazole (white) and PyMT (green), and fraction of PyMT-positive cells in normoxic and hypoxic areas. The area outlined

corresponds to the hypoxic, pimonidazole-positive section, arrowheads point to PyMT-negative cells. Scale bar, $25\mu\text{m}$. The bar chart inset illustrates the relative number of PyMT-positive cells in normoxic and hypoxic areas (grey and red, respectively; $n = 19$). **f**, Ki67-positive cells in PyMT tumours: representative image of staining for DNA (propidium iodide, blue), Ki67 (red) and pimonidazole (green). Scalebar, $50\mu\text{m}$. The bar chart inset illustrates the quantification of Ki67-positive cells in normoxic and hypoxic areas (grey and red, respectively) across 6 tumours, analysing 3 fields of view with over 150 cells per field of view. **g**, CD45-positive cells in PyMT tumours: representative image of staining for DNA (propidium iodide, blue), 5hmC (red), pimonidazole (green) and CD45 (white). Scale bar, $100\mu\text{m}$. The bar chart inset illustrates the quantification of CD45-positive cells in normoxic and hypoxic areas (white and red, respectively) of 11 tumours, capturing on average $\sim 2,500$ nuclei per analysis. *** $P < 0.001$ in (**a**) by Fisher's exact test, significance relative to all genes.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Manipulation of tumour oxygenation in mouse breast tumours, and effects on 5hmC, TSG hypermethylation and confounders.

a, Plasma sFlk1 concentrations at the indicated times after hydrodynamic injection with an empty ($n = 7$) or sFlk1-overexpression plasmid ($n = 5$) (grey and red, respectively). **b, c** Quantification of tumour vessel number (**b**) and hypoxic areas (**c**) of tumours from transgenic MMTV-PyMT mice, hydrodynamically injected with an empty or sFlk1-overexpression plasmid, with representative images of blood vessels stained for CD31 (**b**) and hypoxic areas stained for pimonidazole adducts (**c**). Scale bar, 100 μm . **d**, Changes in RNA expression of hypoxia marker genes that are known to be downregulated (*Mrc1*) or upregulated (*Bnip3*, *Car9*, *Ddit4*) in hypoxic conditions. **e**, 5hmC levels (y axis) across mouse chromosome 18 (x axis) in 400 kb bins, with the location of RefSeq genes (middle), and differences in 5hmC levels (lower). 5hmC levels were determined using shallow TAB-seq, and chromosome 18 was selected because it has large stretches of gene deserts that illustrate the 5hmC depletion in these areas ($n = 3$). 5hmC levels decrease by $12.4\% \pm 3.5$ after sFlk1 overexpression, although technical limitations of TAB-seq (incomplete 5hmC protection or bisulfite-conversion) may partially obscure the magnitude of effects. **f**, Hypermethylation in tumours developing in 12-week-old mice receiving hydrodynamic injection with an empty ($n = 19$) or sFlk1-overexpressing plasmid ($n = 24$) 3 weeks earlier. DNA was bisulfite-converted, PCR-amplified for the indicated oncogenes, and sequenced to a depth of $\sim 500\times$. Plotted are z -scores of hypermethylation (y -axis, exponential), relative to the more normoxic tumours (that is, empty). The dotted line represents the threshold at 5% FDR, and bold darker dots the tumours displaying significantly increased hypermethylation events. **g–j**, Relative

weights of tumours from tg(MMTV-PyMT) mice, hydrodynamically injected with an empty (grey, $n = 19$) or sFlk1-overexpression plasmid (red, $n = 24$) (**g**), and corresponding RNA expression of *Ptprc* (the gene encoding CD45, $n = 5$) (**h**), of *Tet* enzymes (**i**, $n = 15$ for empty plasmid, $n = 12$ for sFlk1-overexpressing plasmid) and of cell proliferation markers (**j**, $n = 5$ for each). **k–m**, As in **d–f** but for 16-week old transgenic MMTV-PyMT mice of the indicated genotype. $n = 5$ (**k**), $n = 3$ for *Phd2*^{+/+}; $n = 4$ for *Phd2*^{fl/fl} (**l**) and $n = 9$ (**m**). **n**, As in **d**, but for 16-week-old *Tie2*-Cre;Tg(MMTV-PyMT) mice of the indicated genotypes ($n = 5$). **o**, DNA was extracted from 17 breast tumours developing in *Tie2*-Cre;*Phd2*^{fl/WT}; Tg(MMTV-PyMT) mice (blue) and 13 breast tumours developing in *Tie2*-Cre;*Phd2*^{WT/WT};Tg(MMTV-PyMT) mice (grey), all 16 weeks old. DNA was bisulfite-converted, PCR-amplified for the indicated TSGs (left) and oncogenes (middle) and sequenced to a depth of $>500\times$. Plotted are z -scores of hypermethylation (y axis, exponential), relative to the more normoxic, *Phd2*^{WT/fl}, tumours. The dotted line represents the threshold for a Bonferroni-adjusted $P < 0.05$, and bold darker dots the tumours displaying significantly increased hypermethylation events. Right, 5hmC levels \pm s.e.m. across a metagene in tumours of 16-week-old mice with the indicated genotype ($n = 3$ for *Phd2*^{fl/fl}; $n = 4$ for *Phd2*^{WT/fl}). **p–u**, Relative weights of tumours from *Phd2*^{+/+};tg(MMTV-PyMT) mice and *Phd2*^{+/+};Tg(MMTV-PyMT) mice ($n = 10$ and 9 resp.) (**p–r**) and from *Tie2*-Cre;*Phd2*^{fl/WT};Tg(MMTV-PyMT) and *Tie2*-Cre;*Phd2*^{WT/WT};Tg(MMTV-PyMT) mice ($n = 17$ and 13, respectively) (**s–u**), and the corresponding RNA expression of cell proliferation markers ($n = 5$, **p, s**), of *Tet* enzymes ($n = 5$, **q, t**) and of *Ptprc* ($n = 5$) (**r, u**). # $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by t -test.

4 Unveröffentlichte Arbeiten

4.1 Quantifizierung von Ethenoaddukten

4.1.1 Prolog

Ethenoaddukte gelten wie zuvor beschrieben als sehr mutagen und werden durch die Reaktion der DNA Basen mit Lipidperoxydationsprodukten wie 4-Hydroxynonenal oder Malondialdehyd gebildet. Diese entstehen durch entzündungsbedingten oxidativen Stress oder die Induktion von Cytochrom P-450 2E1 (CYP2E1). In einer Studie konnte die Bildung von Ethenoaddukten mit der Induktion von CYP2E1 korreliert und so ein direkter Zusammenhang zwischen Alkoholismus oder chronischer Hepatitis und dem Auftreten von Ethenoaddukten hergestellt werden.^[112]

In früheren Arbeiten sind bereits Methoden zur Quantifizierung von Ethenoaddukten beschrieben. Taghizadeh *et al* zeigen eine sensitive Methode zur Quantifizierung verschiedener DNA Schäden. Gekoppelt ist hier eine Anreicherung der Analyten mittels HPLC mit anschließender Analyse via LC/MS-MS mit einem Triple Quadrupol Massenspektrometer.^[116] Für die Quantifizierung der Ethenoaddukte 1,*N*⁶-etheno-2'-desoxyadenosin und 3,*N*⁴-etheno-2'-desoxycytidin aus Zellen oder Geweben waren mindestens 10-20 µg beziehungsweise 50 µg isolierte DNA und ein Anreicherungsschritt mittels HPLC nötig. Für die Quantifizierung von Ethenoaddukten in geringeren Mengen an DNA wäre eine noch sensitivere Methode notwendig.

4.1.2 Ergebnisse und Diskussion

Eine sensitive LC/MS-MS Methode zur Quantifizierung der Ethenoaddukte 1,*N*⁶-etheno-2'-desoxyadenosin (ϵ dA) und 3,*N*⁴-etheno-2'-desoxycytidin (ϵ dC) sollte in Kooperation mit der Gruppe von Prof. Dr. med. Schlegel aus der Neuropathologie der TU München entwickelt werden. Die chromatographische Auftrennung sollte ohne vorherige Anreicherung mit einer UHPLC und die anschließende Analyse mit einem Triple Quadrupol Massenspektrometer erfolgen. Da die Methode auch zur Quantifizierung von Patientenproben genutzt werden sollte, war es wichtig eine möglichst sensitive Methode für geringe Mengen an DNA zu entwickeln.

Die beiden Standards ¹⁵N₅-*N*⁶-etheno-2'-desoxyadenosin (¹⁵N₅- ϵ dA) und ¹⁵N₃-*N*⁴-etheno-2'-desoxycytidin (¹⁵N₃- ϵ dC) (Abb 4.1) wurden vom Kooperationspartner synthetisiert und bereitgestellt.

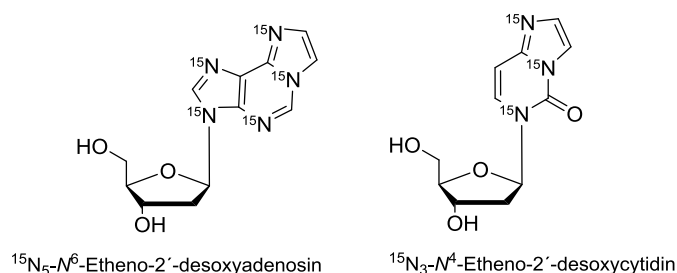
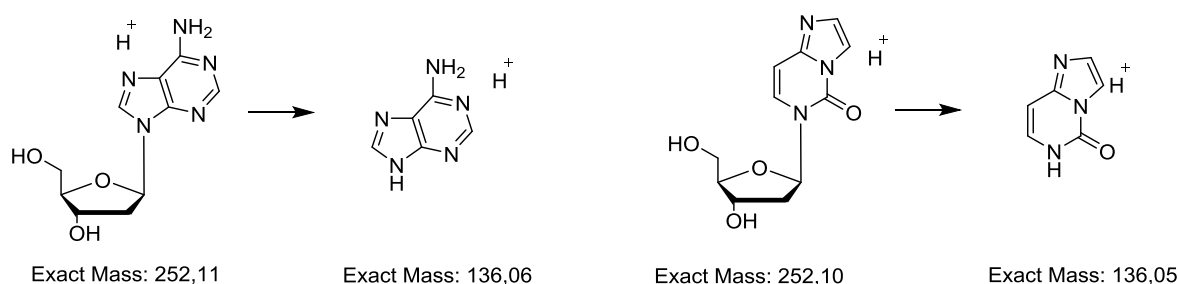
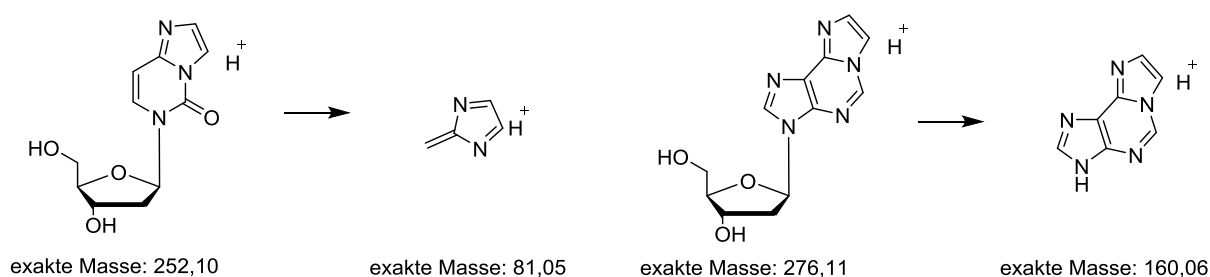


Abb. 4.1: Synthetische Standards zur Quantifizierung der Ethenoaddukte 1,*N*⁶-Etheno-2'-desoxyadenosin (ϵ dA) und 3,*N*⁴-Etheno-2'-desoxycytidin (ϵ dC)

Zunächst wurde eine UHPLC-Methode zu Auftrennung der beiden Ethenoaddukte und zur Abtrennung weiterer DNA Bestandteile entwickelt. Die besondere Herausforderung bestand darin, ¹⁵N₃- ϵ dC von natürlichen 2'-Desoxyadenosin (dA) abzutrennen, da dA dasselbe Fragmentierungsmuster (Abb 4.2) sowie eine ähnliche Retentionszeit beim gewählten Gradienten und Lösungsmittelgemisch aufweisen.

Abb. 4.2: Intensivste Fragmentierungen von 2'-Desoxyadenosin und N^4 -Etheno-2'-desoxycytidin

Um dieses Problem zu lösen wurden zwei Strategien verfolgt: Zum einen wurde der Elutionsgradient dahingehend verändert, dass die beiden Substanzen eine etwas unterschiedliche Retentionszeit aufweisen, zum anderen wurde versucht eine zweite, intensive Fragmentierung für $^{15}\text{N}_3$ - ϵdC zu finden. Dies konnte mit gesteigerter Kollisionsenergie (CE) erreicht werden. Es wurde eine zweite ebenso intensive Fragmentierung entdeckt, welche für dA nicht gefunden werden konnte. Somit waren für beide Ethenoaddukte spezifische Übergänge gefunden, die bis in den attomolaren Bereich (400 amol) detektiert werden konnten. (Abb. 4.3)

Abb. 4.3: Intensivste Fragmentierungen von $^{15}\text{N}_5$ - N^6 -Etheno-2'-desoxyadenosin und $^{15}\text{N}_3$ - N^4 -Etheno-2'-desoxycytidin

Die Zellkulturexperimente sowie auch die Isolation der DNA wurden von *M.Sc. Johanna Donhauser* aus der Gruppe von Prof. Dr. med. Schlegel durchgeführt. In einem Experiment sollte die Wirkung der Aldehyddehydrogenase 1A1 (ALDH1A1) auf die durch Temozolomid (TMZ) induzierten Ethenoaddukte untersucht werden. Temozolomid ist ein oral und intravenös verabreichtes, alkylierendes Zytostatikum. Es kann simultan oder ergänzend in Kombination mit Strahlentherapie angewandt werden.^[159] Die Wirkung beruht auf einer Störung der Tumorzelle in der DNA

Replikation.

Als Nebenwirkung wurden unter anderem Leberschäden beschrieben.^[160] Dies könnte durch Lipidperoxidation hervorgerufen werden und daher mit der Bildung von Ethenoaddukten in Zusammenhang stehen. Im Experiment sollte getestet werden, ob die Aldehyddehydrogenase 1A1 (ALDH1A1) eine Resistenz gegen TMZ vermitteln und somit die Bildung von Ethenoaddukten reduzieren kann.

Es wurden zwei Zelllinien verwendet. Zum einen die ALDH1A1-positive, TMZ resistente Zelllinie LN18 (LN18 wt), zum anderen eine Zelllinie mit einem shRNA Knockdown von ALDH (LN18shALDH). Da durch die RNA-Interferenz kein stabiler *Knockdown* von ALDH erhalten wurde, wurde später zusätzlich der chemische ALDH Inhibitor DEAB (Diethylaminobenzaldehyd) verwendet. Als Kontrolle wurde auch das Lipidperoxidationsprodukt 4-Hydroxynonenal (HNE) verabreicht, von dem bekannt ist, dass es Ethenoaddukte induziert. Als Lösungsmittelkontrollen zu TMZ und HNE dienten die jeweiligen Lösungsmittel Dimethylsulfoxid (DMSO) und Ethanol (EtOH). Die unbehandelte Kontrolle wird mit KO für Kontrolle abgekürzt. Für beide Ethenoaddukte wäre zu erwarten gewesen, dass unter HNE Behandlung wie auch unter TMZ Behandlung der ALDH inhibierten Zellen mehr Ethenoaddukte entstehen als in den unbehandelten Kontrollen. Des Weiteren sollte dieser Effekt in Wildtyp-zellen schwächer oder sogar ganz aufgehoben sein.

In einer ersten Messung konnten mit 20 µg DNA ohne vorherige HPLC Anreicherung beide Ethenoaddukte zuverlässig quantifiziert werden. (Abb. 4.4)

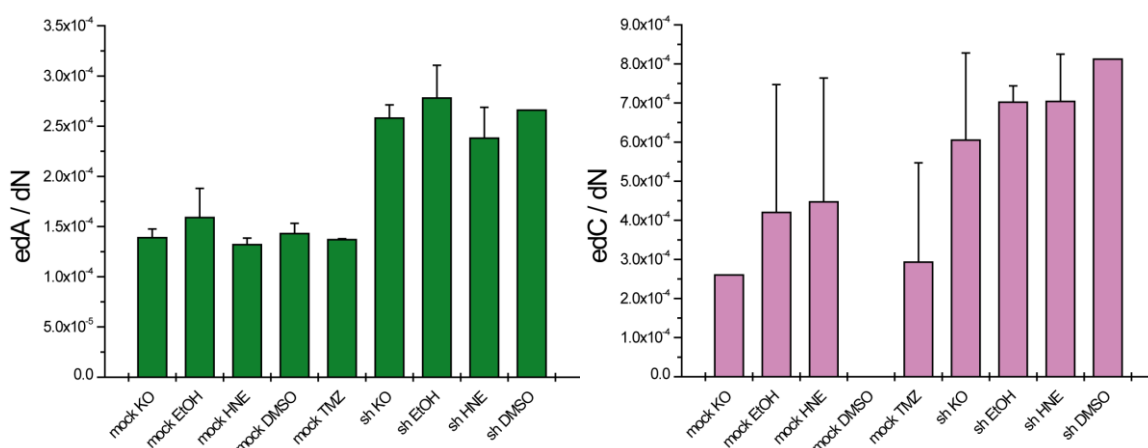


Abb. 4.4: Absolute Werte an 1,N⁶-Ethno-2'-desoxyadenosin (grün) und 3,N⁴-Ethno-2'-desoxycytidin (violett) mit den Kontrollen KO (unbehandelte Kontrolle), EtOH (Lösungsmittelkontrolle zu HNE) und DMSO (Lösungsmittelkontrolle zu TMZ) und den Proben die mit HNE und TMZ behandelt wurden. Mock bedeutet

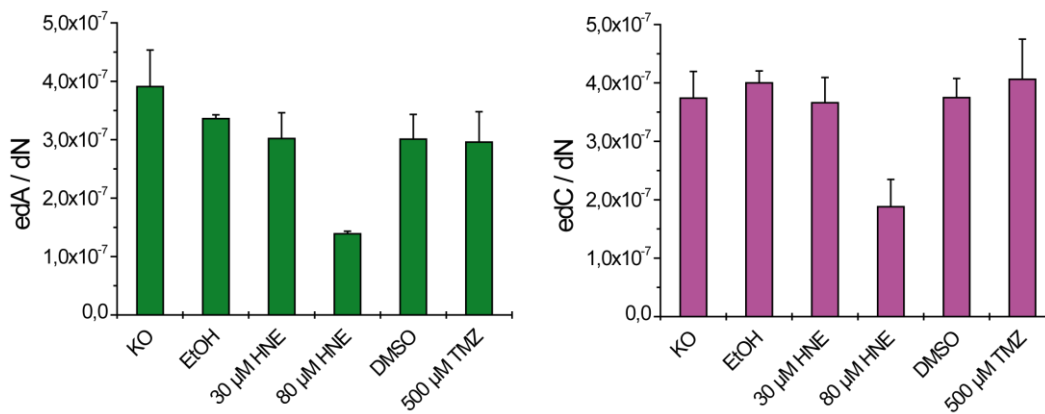
Messung im Wildtyp LN18 (LN18 wt), sh ist die Zelllinie mit ALDH Knockdown (LN18shALDH). Für mock DMSO konnten für edC keine Werte ermittelt werden.

Man erkennt, dass sowohl die Level von ϵ dA als auch von ϵ dC im Knockdown von ALDH1A1 (LN18shALDH) erhöht sind. Es ist jedoch auch ersichtlich, dass dies in allen Proben der Fall ist. Eigentlich sollte nur der Level der Proben die mit HNE und TMZ behandelt wurden, erhöht sein. Die Lösungsmittelkontrollen sollten im ALDH1A1 Knockdown gleich bleiben. Somit scheint entweder schon das Lösungsmittel einen Effekt zu haben oder der Knockdown hat nicht funktioniert.

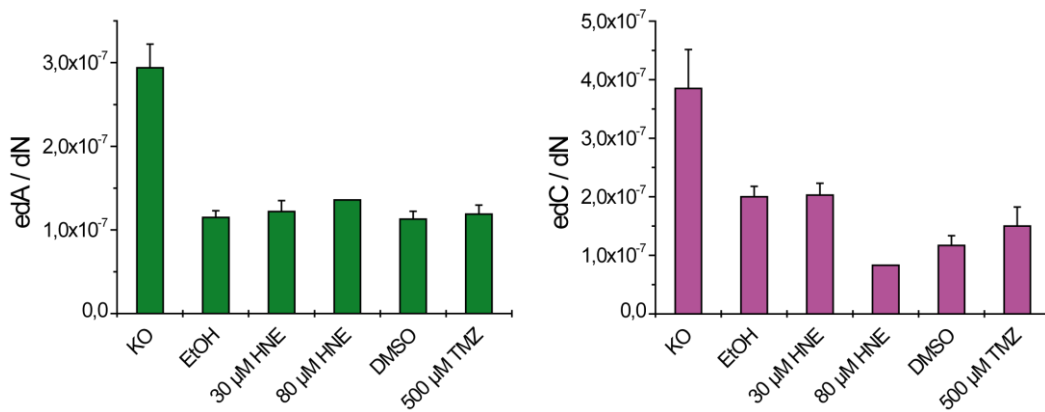
Um zu überprüfen, ob es am Knockdown lag, wurde das Experiment wiederholt und diesmal zusätzlich der chemische ALDH1A1 Inhibitor DEAB (Diethylaminobenzaldehyd) verwendet. Dieser sollte die ALDH1A1-Level zusätzlich reduzieren und so den Effekt verstärken.

Im zweiten Experiment scheint der Knockdown jedoch gar nicht funktioniert zu haben. Man sieht sowohl für den Knockdown selbst (Abb. 4.5 B), als auch unter Einsatz des Inhibitors (Abb. 4.5 C) für beide Ethenoaddukte durchgehen geringere Level als im Wildtyp. (Abb. 4.5 A)

A



B



C

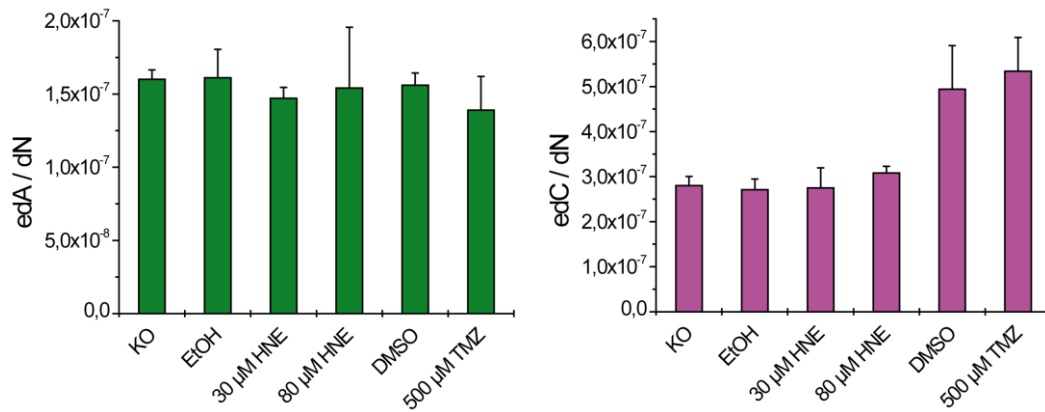


Abb. 4.5: Absolute Werte an 1, N^6 -Etheno-2'-desoxyadenosin (grün) und 3, N^4 -Etheno-2'-desoxycytidin (violett) mit den Kontrollen KO (unbehandelte Kontrolle), EtOH (Lösungsmittelkontrolle zu HNE) und DMSO (Lösungsmittelkontrolle zu TMZ) und den Proben die mit 30 und 80 μM HNE und 500 μM TMZ behandelt wurden. Oben sind die Messungen im Wildtyp LN18 (LN18 wt) (A) darunter in der Zelllinie mit ALDH Knockdown (LN18shALDH) (B) und LN18shALDH unter zusätzlicher Verwendung von DEAB (C).

Auffällig ist, dass unter Einsatz des ALDH1A1-Inhibitor (Abb. 4.5 C) für ϵ dC eine leichte Zunahme sowohl für TMZ als auch für dessen Kontrolle sichtbar ist. Hier liegen die Werte über dem Level an ϵ dC im Wildtyp. Für ϵ dA ist dies allerdings nicht zu beobachten. Um ein klareres Bild von der Wirkung von TMZ auf Ethenoaddukte und den Einfluss der ALDH1A1 in diesem Zusammenhang zu erhalten, müsste das Experiment mit einem stabilen Knockout und möglicherweise mit noch größeren Mengen an HNE und TMZ wiederholt werden. Das Kooperationsprojekt wurde an dieser Stelle jedoch vom Kooperationspartner aus Personalmangel eingestellt.

4.1.4 Zusammenfassung und Projektbeiträge

Im Rahmen dieses Projektes wurde eine Methode zur sensitiveren Quantifizierung der beiden Ethenoaddukte 1,*N*⁶-Etheno-2'-desoxyadenosin und 3,*N*⁴-Etheno-2'-desoxycytidin ohne vorher benötigte Anreicherung entwickelt. Auf Grund der hohen Sensitivität der Methode können selbst in geringen Mengen DNA beide Ethenoaddukte zuverlässig quantifiziert werden. Somit könnten auch DNA Proben vermessen werden, die aus dem Blut oder den Zellen von Patienten gewonnen werden. Die Methode bietet somit Vorteile gegenüber der Methode von Taghizadeh *et al.*^[116]

Das Experiment müsste auf Grund des mangelhaften Knockdowns von ALDH1A1 wiederholt werden, um die Frage nach dem Einfluss von TMZ auf Ethenoaddukte und die Rolle von ALDH1A1 zu klären.

In Zusammenarbeit mit *M.Sc. Johanna Donhauser* entwickelte ich einen geeigneten enzymatischen Verdau der isolierten DNA zu Quantifizierung der Ethenoaddukte. Die Experimente an den Zellen sowie die nachfolgende Isolation der DNA wurden von *M.Sc. Johanna Donhauser* aus der Gruppe von Prof. Dr. med. Schlegel durchgeführt. Die Synthese der isotoplegelabelten Standards ¹⁵N₅-*N*⁶-Etheno-2'-desoxyadenosin (¹⁵N₅-εdA) und ¹⁵N₃-*N*⁴-Etheno-2'-desoxycytidin (¹⁵N₃-εdC) wurde ebenfalls in Kooperation mit der Gruppe von Prof. Dr. med. Schlegel von *M. Sc. Vanessa Röhrig* durchgeführt.

Ich entwickelte und optimierte eine LC/MS-MS Methode zur sensitiven Quantifizierung von Ethenoaddukten und Qualifizierte die Methode. Die Messungen, Auswertung und Darstellung der Daten wurde ebenfalls von mir vorgenommen.

4.1.5 Material und Methoden

DNA Isolation

Die Isolation der DNA erfolgte nach dem unter 3.3 beschriebenen Protokoll.

Verdau

Der Verdau der genomischen DNA (10-20 µg) erfolgte in 50 µL ddH₂O. Die DNA wurde in 35 µL ddH₂O vorgelegt. Für den ersten Verdauschnitt wurden 7.5 µL wässrige Lösung zugegeben. Diese enthielt 480 µM ZnSO₄, 50 U Nuklease P1 (Sigma-Aldrich), 5 U Antarctic Phosphatase (NEB) und eine spezifische Menge an markierten internen Standards. Die Mischung wurde bei 37°C für 16 h inkubiert. Für den zweiten Verdauschnitt wurden ebenfalls 7.5 µL wässrige Lösung zugegeben. Diese enthielt 520 µM Na₂-EDTA-Lösung und 0.2 U Snake Venom Phosphodiesterase (*Crotalus adamanteus*, USB Corporation) und wurde bei 37°C für 3h inkubiert. Die Proben wurden bis zum Tag der Analyse bei -20°C aufbewahrt. Vor der UHPLC-MS/MS Analyse wurden die Proben mit AcroPrep Advance 96 Filterplatten 0.2 µm Supor (Pall Life Sciences) filtriert.

LC-MS/MS-Methode

Die quantitative LC-ESI-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte mit einem *Agilent* 1290 UHPLC-System, welches über einen UV Detektor verfügt und mit einem *Agilent* 6490 Triple-Quadrupol Massenspektrometer verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8-Säule (*Agilent*, 2.7 µm, 2.1 mm × 150 mm) bei 30°C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0.0085% (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0.35 ml/min: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 7.2 min; 5 → 80% MeCN; 7.2 → 10.5 min; 80% MeCN; 10.5 → 11.3 min; 80 → 0% MeCN; 11.3 → 13 min; 0% MeCN.

Die Messungen erfolgten mit der unter 3.3 bereits beschriebenen Isotopenverdünnungsmethode und basiert auf früher veröffentlichten Arbeiten.^[85, 161, 162] Die Quellparameter waren: Gastemperatur 50 °C, Gasfluss 15 l/min (N₂), Nebulizer 30 psi, Sheath gas Heizer 275 °C, *Sheath gas* Fluss 11 l/min (N₂), *capillary voltage* 2,500 V im positiven Modus, *capillary voltage* -2,250 V im negativen Modus und *nozzle voltage* 500 V. Die Fragmentorspannung betrug 380 V. Delta EMV betrug 500 (positiver Modus) und 800 (negativer Modus). Die für die jeweiligen Analyten spezifischen Parameter sind in Tabelle 4.1 zusammengefasst.

Tabelle 4.1: Substanz abhängige LC/MS-MS Parameter zur Quantifizierung von Ethenoaddukten in genomischer DNA

Verbindung	Precursor Ion (<i>m/z</i>)	MS1 Auflösung	Produkt Ion (<i>m/z</i>)	MS2 Auflösung	Dwell time [ms]	CE (V)	CAV (V)	Polarität
dC	228.10	Unit	112.05	Unit	40	1	0	positiv
[¹⁵ N ₃]-edC	255.10	Wide	83.05	Wide	50	53	5	positiv
edC	252.11	Wide	81.05	Wide	50	53	5	positiv
[¹⁵ N ₅]-edA	281.10	Wide	165.05	Wide	50	8	5	positiv
edA	276.11	Wide	160.06	Wide	50	8	5	positiv
[¹⁵ N ₅]-8-oxo-G	289.08	Wide	173.04	Wide	40	9	7	positiv
8-oxo-G	284.10	Wide	168.05	Wide	40	9	7	positiv
dT	243.10	Unit	127.05	Unit	40	40	3	positiv
dA	252.11	Unit	136.06	Unit	40	45	1	positiv

Validierung der LC-MS Methode

Die entwickelte Methode wurde in Übereinstimmung mit der Anleitung der FDA für bioanalytische Methodenvalidierung auf Linearität, Präzision und Genauigkeit überprüft. Die Validierung der LC-ESI-MS/MS Quantifizierungsmethode erfolgte auf verschiedene Arten (Kalibrierung und Proben zu Qualitätskontrolle) an unterschiedlichen Tagen. Jeder Kalibrierungsstandard wurde vier bis sechsmal analysiert und die Validierungsexperimente durch Einsatz von Blindproben hinsichtlich ihrer Selektivität und Spezifität untersucht.

Eichgeraden:

Die für die Quantifizierung notwendigen Eichgeraden der markierten und entsprechenden unmarkierten synthetisierten Nukleoside wurden bei 6 bzw. 4 unterschiedlichen Konzentrationen aufgenommen. Die lineare Regression wurde unter Verwendung von Origin 6.0 (Microcal) durchgeführt. Hierfür wurde das Verhältnis der Flächen (A/A^*) unter dem Signal für das unmarkierte Nukleosid durch den internen Standard (*) gegen das Verhältnis der Stoffmengen (n/n^*) des unmarkierten Nukleosids durch den internen Standard aufgetragen. Für jede Konzentration wurde ein Mittelwert aus drei unabhängigen Messungen gebildet. (Abb. 4.6).

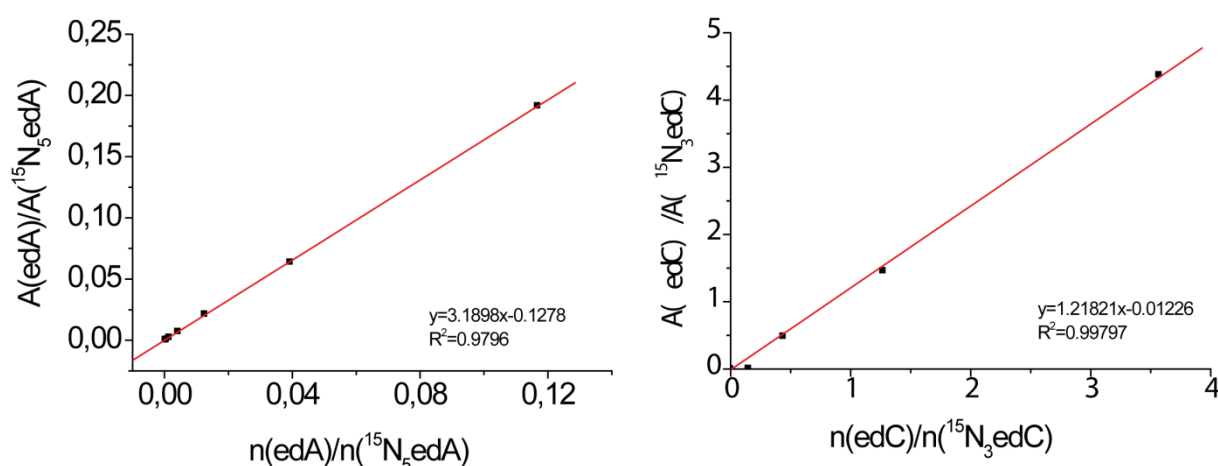


Abb. 4.6: Masseneichgeraden für die Nukleoside edA und edC

Vorausgesetzt wurden eine Genauigkeit von 80-120% sowie eine Präzision von weniger als 20 RSD. Die Genauigkeit wurde überprüft durch Berechnung der Stoffmengen durch das mit Hilfe der Eichgerade erhaltene A/A^* Verhältnis der Analyten. Die Genauigkeit wurde definiert als das Verhältnis aus eingesetzter Menge an Substanz durch die berechnete Menge an Substanz in Prozent. Diese Genauigkeit sollte zwischen 80 und 120% liegen.

Für die Präzision mussten bei technischen Replikaten die Standardabweichung (RSD) unter 20% liegen. Das untere Quantifizierungslimit (*lower limit of quantification*, LLOQ) wurde als niedrigste Konzentration definiert die grade noch die Anforderungen der Genauigkeit und Präzision erfüllt.

Auswertung

Die Auswertung erfolgte mit dem Programm Mass Hunter (Agilent) und Excel (Microsoft), die Darstellung der Daten erfolgte mit Origin und Adobe Illustrator.

4.2 Decarboxylierungsstudien mit D_2 - $^{15}N_2$ -dC

4.2.1 Prolog

Wie bereits unter 1.1.1.7 diskutiert gibt es Argumente, die für eine aktive DNA Demethylierung über eine Decarboxylierung von 5cadC oder eine Deformylierung von 5fdC sprechen.^[17, 53, 71, 72, 84, 86-88, 163] Seit Langem wird in der Carell Gruppe daher nach einem geeigneten Assay gesucht um diese Fragestellung näher zu beleuchten. Mit Hilfe eines an C5 und C6 der Nukleobase doppelt deuterierten und weiterhin doppelt ^{15}N -markierten Cytidines kann sich diesem Problem angenähert werden. Durch die Massenzunahme von 4 Einheiten (dC+4) kann das D_2 - $^{15}N_2$ -dC massenspektrometrisch gut von natürlichem dC unterschieden werden. Zudem erlaubt Die $^{15}N_2$ -Markierung die Unterscheidung zwischen BER und einem C-C-Bindungsbruch. Die doppelte D-Markierung ist notwendig, um zu entscheiden, ob ein Cytidin der gefütterten Version, oder der methylierten/demethylierten Version entspricht, da das Produkt der Demethylierung identisch zum Edukt ist. Somit dient die vierfachen Markierung hier der Aufklärung des genauen Ablaufs der Demethylierung (Abb.4.7).

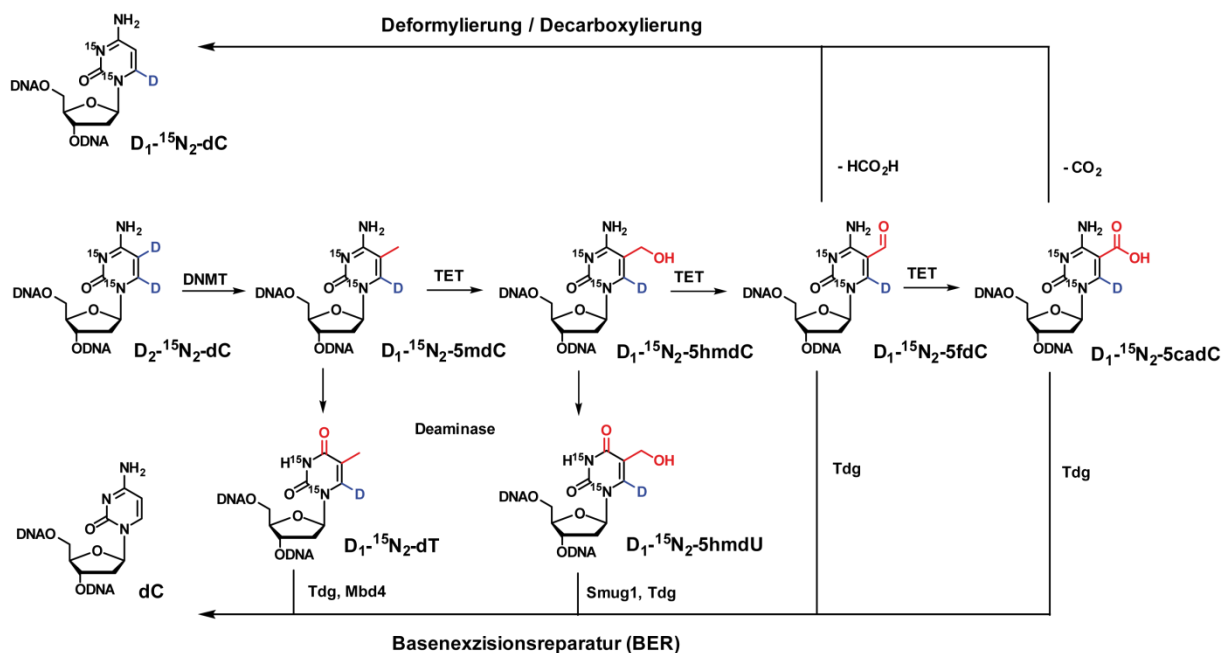


Abb. 4.7: Die möglichen Wege der aktiven DNA-Demethylierung über Basenexzisionsreparatur oder über einen C-C-Bindungsbruch lassen sich durch Verfolgung der Isotopenmarkierung unterscheiden.

Nach einer Methylierung von $D_2\text{-}^{15}\text{N}_2\text{-dC}$ durch DNA-Methyltransferasen entsteht $D_1\text{-}^{15}\text{N}_2\text{-5mdC}$, welches durch die Tet Enzyme zu $D_1\text{-}^{15}\text{N}_2\text{-5hmdC}$ und weiter zu $D_1\text{-}^{15}\text{N}_2\text{-5fdC}$ und $D_1\text{-}^{15}\text{N}_2\text{-5cadC}$ oxidiert werden kann. Durch eine Desaminierung von $D_1\text{-}^{15}\text{N}_2\text{-5mdC}$ und $D_1\text{-}^{15}\text{N}_2\text{-5hmdC}$ können zudem $D_1\text{-}^{15}\text{N}_2\text{-dT}$ und $D_1\text{-}^{15}\text{N}_2\text{-5hmdU}$ entstehen. Anschließend können die oxidierten oder desaminierten Nukleoside durch eine Glykosylase ausgeschnitten über Basenexzisionsreparatur durch ein unmarkiertes dC ersetzt werden.^[99] Auch das direkte Entfernen von $D_1\text{-}^{15}\text{N}_2\text{-5fdC}$ oder $D_1\text{-}^{15}\text{N}_2\text{-5cadC}$ durch eine Deformylierung von $D_1\text{-}^{15}\text{N}_2\text{-5fdC}$ oder eine Decarboxylierung von $D_1\text{-}^{15}\text{N}_2\text{-5cadC}$ wäre denkbar. Hier müsste lediglich die Bindung an C5 gebrochen werden, wodurch die einfach deuterierte Spezies $D_1\text{-}^{15}\text{N}_2\text{-dC}$ entstehen würde. Diese könnte anschließend auch wieder methyliert werden, weswegen im Anschluss nicht nur nach $D_1\text{-}^{15}\text{N}_2\text{-dC}$, sondern ebenso nach $D_1\text{-}^{15}\text{N}_2\text{-5mdC}$ als Produkt der aktiven Demethylierung über Deformylierung und Decarboxylierung gesucht werden soll.

4.2.2 Ergebnisse und Diskussion

Für die Messung und Quantifizierung der deuterierten Analyte D_2 - $^{15}N_2$ -dC und D_1 - $^{15}N_2$ -dC musste die Standard-Messmethode verändert werden. Zunächst wurde nach sensitiven Übergängen für beide Analyte gesucht. Es konnten spezifische Fragmentierungen gefunden werden. (Abb. 4.8)

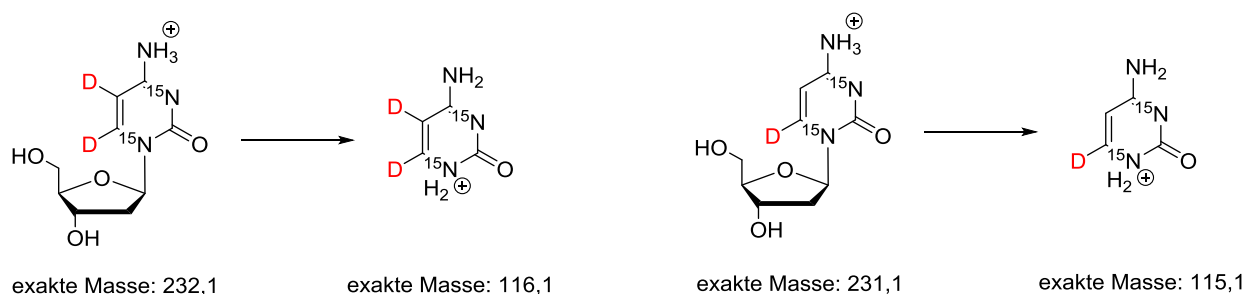


Abb. 4.8: Spezifische Fragmentierungen für D_2 - $^{15}N_2$ -dC und D_1 - $^{15}N_2$ -dC

Der für die Fütterung benötigte, hoch isotopenangereicherte, D_2 - $^{15}N_2$ -dC Standard wurde von Arne Schröder und René Rahimoff synthetisiert und zur Verfügung gestellt. Die Fütterung der verschiedenen Zelltypen mit diesem Nukleosid wurde von Fabio Spada vorgenommen.

Zunächst wurde überprüft, welche Isotopenanreicherung von D_2 - $^{15}N_2$ -dC bei der Synthese erzielt wurde. Mit Hilfe von UHPLC-MS/MS konnte die Isotopenanreinheit auf 95% bestimmt werden. Mit diesem D_2 - $^{15}N_2$ -dC Analyten wurden die ersten Fütterungsexperimenten in embryonalen Stammzellen der Maus (mESC) vorgenommen. In einem ersten Versuch sollte die Fütterung von 100 μ M D_2 - $^{15}N_2$ -dC in TDG Wildtyp (TDG+/-) mESC und TDG *Knockout* (TDG -/-) mESC miteinander verglichen werden. Die Intention war dabei, das sich die Menge an neu gebildetem D_1 - $^{15}N_2$ -dC in den TDG-/- Zellen erhöht, da der zweite mögliche Demethylierungsweg über TDG und BER durch das Fehlen von TDG nicht mehr möglich ist.

Hierbei zeigte sich, dass durch die Fütterung von D_2 - $^{15}N_2$ -dC an mESC ein Markierungsgrad von 39% erreicht werden konnte. Zudem konnte nachvollzogen werden, dass sich das markierte D_2 - $^{15}N_2$ -dC in Bezug auf die weitere Methylierung und Oxidation exakt wie das unmarkierte dC verhält. Es wurde zu 3.9% zu D_1 - $^{15}N_2$ -5mdC methyliert, lag zu 0.7% hydroxymethyliert und zu etwa 0.02% formyliert vor, was auch

den Verhältnissen der unmarkierten Formen entspricht. (Abb. 4.9)

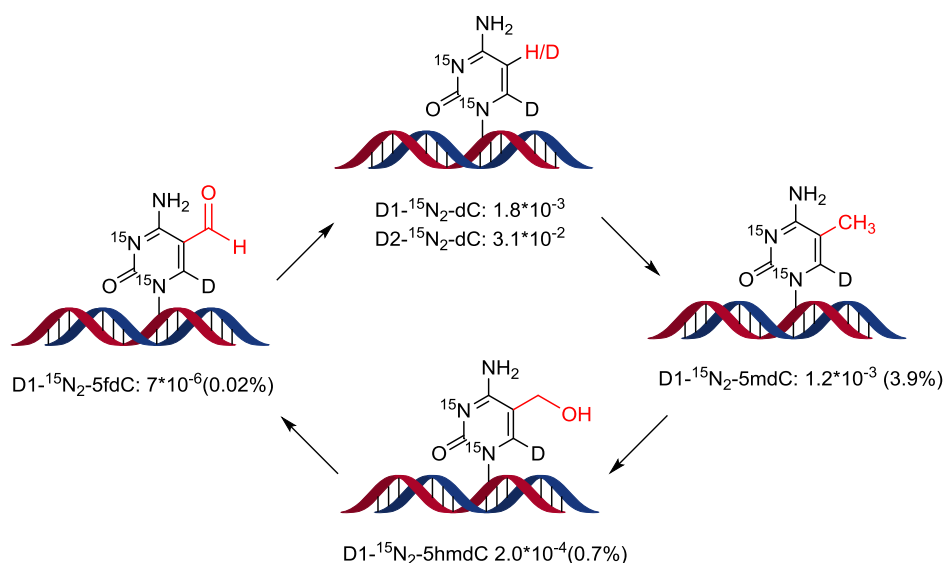


Abb. 4.9: Absolute Quantifizierung der Menge an $\text{D}_2\text{-}^{15}\text{N}_2\text{-dC}$ sowie der Verunreinigung an $\text{D}_1\text{-}^{15}\text{N}_2\text{-dC}$ und der methylierten und oxidierten Spezies in der DNA.

An dieser Stelle wurde jedoch auch festgestellt, dass ein Isotopenanteil von 95% auf Grund des relativ hohen Hintergrundes an $\text{D}_1\text{-}^{15}\text{N}_2\text{-dC}$ im Ausgangsmaterial für die Untersuchung der Decarboxylierung, welche sich in einem wesentlich kleineren Größenmaßstab abspielen sollte, nicht ausreicht. Aus diesem Grund wurde die Synthese überarbeitet und es konnte schließlich von René Rahimoff eine Reinheit von 97% erreicht werden. Aufgrund der Isotopenreinheit der Edukte, welche für $^{15}\text{N}_2\text{-Harnstoff}$ nur 98% beträgt und aufgrund eines nicht vollständigen Austausches des C5-H mit Deuterium konnte diese Reinheit nicht weiter erhöht werden.

Somit war nun die bestmögliche Isotopenreinheit erreicht und es wurde erneut eine Fütterung in TDG +/- und TDG -/- mESC vorgenommen. Diesmal wurde zudem über einen Zeitraum von 3 Tagen gefüttert. Dadurch sollte es zu einer maximalen Anreicherung von $\text{D}_1\text{-}^{15}\text{N}_2\text{-dC}$ im Genom kommen, sodass dieses besser über dem immer noch relativ hohen chemischen Hintergrund von etwa 3% quantifiziert werden kann. Zunächst wurden die Mengen an $\text{D}_1\text{-}^{15}\text{N}_2\text{-mdC}$, $\text{D}_1\text{-}^{15}\text{N}_2\text{-hmdC}$ und $\text{D}_1\text{-}^{15}\text{N}_2\text{-fdC}$ nach einigen Stunden (Tag 0) und 72 h später (Tag 3) mit Hilfe der absoluten Quantifizierung von 5mdC, 5hmdC und 5fdC ermittelt. (Abb. 4.10)

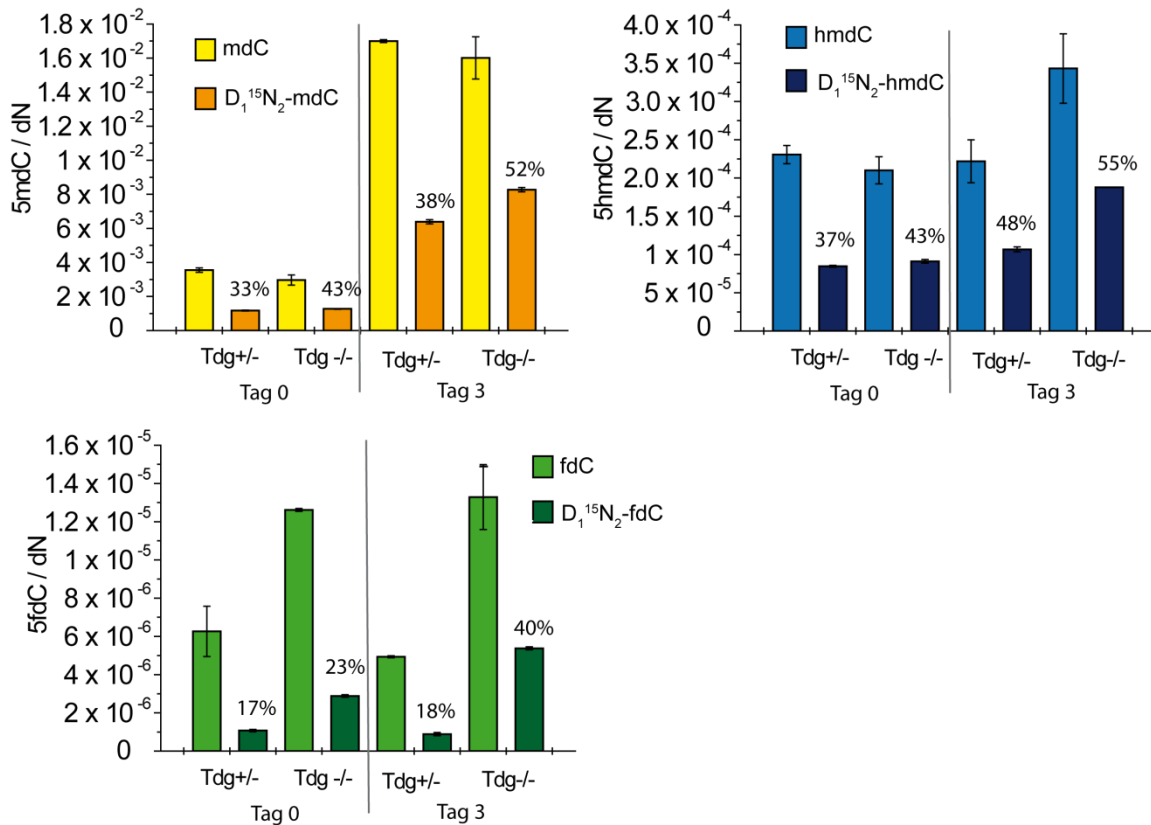


Abb. 4.10: Absolut quantifizierte Mengen an mdC, hmdC und fdC sowie über den prozentualen Anteil berechnete Mengen der deuterierten Spezies $D_1^{15}N_2$ -mdC, $D_1^{15}N_2$ -hmdC und $D_1^{15}N_2$ -fdC an den unmarkierten Modifikationen jeweils zum Zeitpunkt Tag 0 und Tag 3 in TDG +/- Zellen sowie in TDG-/- Zellen. Prozentangaben beziehen sich auf den Anteil an modifizierter Spezies im Vergleich zur unmodifizierten Spezies.

Bei einer Betrachtung der Gesamtmengen (markierte und unmarkierte Nukleoside) fällt zunächst auf, dass an Tag 3 der Ausdifferenzierung deutlich mehr 5mdC vorhanden ist als an Tag 0. Dies ist sowohl im TDG+/- Zellen als auch in den TDG-/- Zellen der Fall. Die 5hmdC-Level bleiben bei den TDG+/- Zellen etwa konstant, nehmen in den TDG-/- Zellen zum Tag 3 allerdings zu. Die 5fdC Mengen wiederum sind an Tag 0 und Tag 3 in etwa gleich, es findet sich jedoch sowohl an Tag 0 als auch an Tag 3 eine Verdopplung der 5fdC Mengen, wenn TDG ausgeschaltet ist.

Betrachtet man nun die isotonenmarkierten Verbindungen, so erkennt man, dass es in den TDG-/- Zellen jeweils zu einer leichten Anreicherung der isotonenmarkierten Spezies im Vergleich zur natürlichen Spezies kommt. Im Falle des $D_1^{15}N_2$ -mdC ergibt sich eine Anreicherung im TDG-/- von 10% an Tag 0 und 14% an Tag 3. Für $D_1^{15}N_2$ -hmdC ist dies ebenfalls zutreffend. Auch hier sieht man, wenn auch etwas weniger ausgeprägt eine Anreicherung an $D_1^{15}N_2$ -hmdC pro 5hmdC um 6% an Tag 0 und 7% an

Tag 3 im TDG-/- . Bei $D_1\text{-}^{15}\text{N}_2\text{-fdC}$ sieht man diese prozentuale Zunahme im Vergleich zur unmarkierten Spezies ebenfalls. Hier ist allerdings auffällig, dass sich die Zunahme an $D_1\text{-}^{15}\text{N}_2\text{-fdC}$ pro 5fdC im TDG-/- an Tag 0, welche 6% beträgt, sehr stark von der prozentualen Zunahme an Tag 3 unterscheidet. Hier findet man 22% mehr $D_1\text{-}^{15}\text{N}_2\text{-fdC}$ pro 5fdC im TDG-/- als im TDG+/- . Somit scheint sich das markierte $D_1\text{-}^{15}\text{N}_2\text{-fdC}$ im TDG-/- über die Tage in besonderer Weise anzureichern.

In den TDG-/- Zellen kommt es somit zu einer Anreicherung der isotoopenmarkierten Spezies. Dies könnte zum einen bedeuten, dass TDG die markierte Spezies bevorzugt entfernt, wodurch es bei einem Fehlen von TDG zu einer vermehrten Anreicherung kommen würde. Allerdings sollte der Isotopeneffekt keinen großen Einfluss auf die Prozessierung durch TDG haben. Zum anderen könnte dies darauf hindeuten, dass die markierten $D_1\text{-}^{15}\text{N}_2\text{-fdCs}$ eher an Positionen im Genom sitzen, welche zu den von Balasubramian *et al.* postulierten stabilen 5fdCs zählen und somit weniger *turnover* zeigen.^[164]

Anschließend wurde die Menge an dC, $D_1\text{-}^{15}\text{N}_2\text{-dC}$ und $D_2\text{-}^{15}\text{N}_2\text{-dC}$ zum Zeitpunkt Tag 0 und nach 3 Tagen jeweils im TDG+/- sowie im TDG Knockout (TDG-/-) absolut quantifiziert und die relativen Anteile bestimmt. (Abb. 4.11)

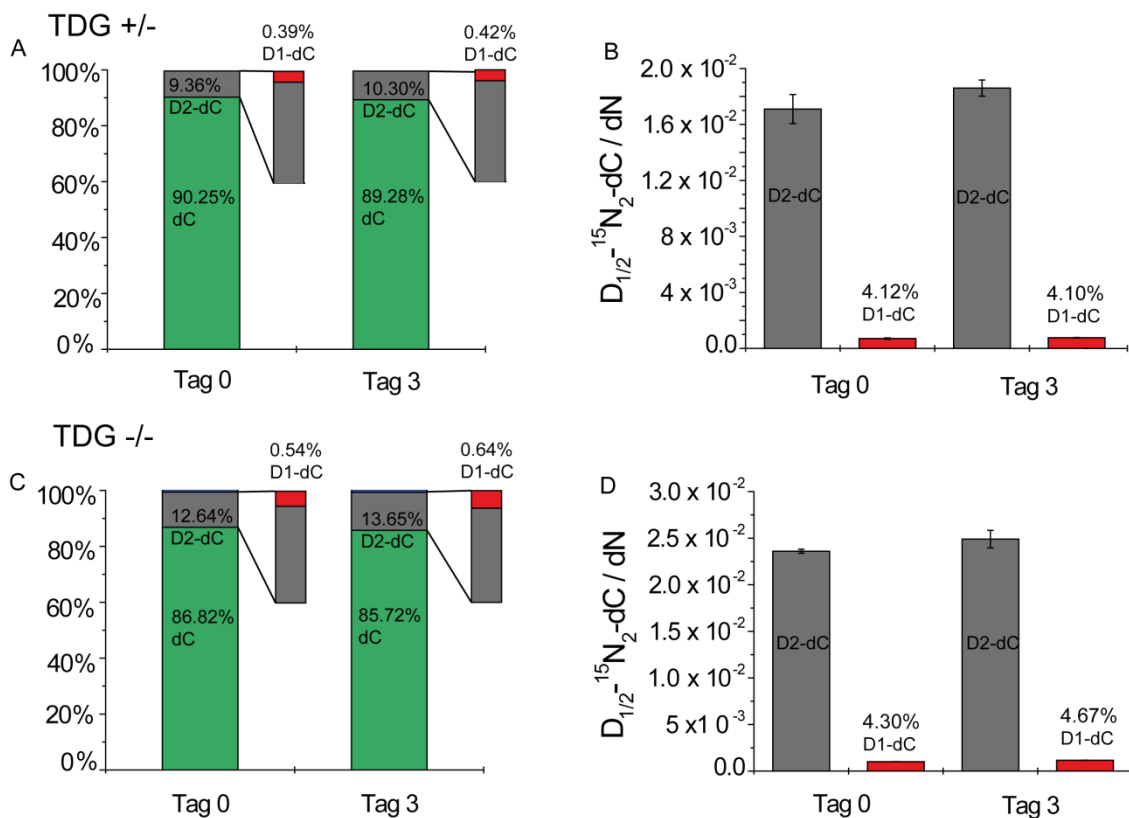


Abb. 4.11: A)+C) Relative Darstellung der Menge an dC, $D_2^{-15}N_2$ -dC und $D_1^{-15}N_2$ -dC in Prozent zum Zeitpunkt Tag 0 und Tag 3. B)+D) Absolute Quantifizierung der Mengen an $D_2^{-15}N_2$ -dC und $D_1^{-15}N_2$ -dC jeweils zum Zeitpunkt Tag 0 und Tag 3. Berechnung der prozentualen Menge an $D_1^{-15}N_2$ -dC im Vergleich zu $D_2^{-15}N_2$ -dC zum Zeitpunkt 0 und nach 3 Tagen.

In Abbildung 4.11 A und C sind die relativen Mengen an dC, $D_2^{-15}N_2$ -dC und $D_1^{-15}N_2$ -dC als prozentualer Anteil aller dC-Spezies dargestellt. In dieser Darstellungsweise erkennt man, dass die Anteile an unmarkiertem dC nach 3 Tagen sowohl im TDG+/- als auch im TDG-/- etwas abnehmen. Die Anteile an $D_2^{-15}N_2$ -dC nehmen in beiden Zelllinien leicht zu. Der Anteil an $D_1^{-15}N_2$ -dC bleibt in den TDG+/- Zellen unter Berücksichtigung der Standardabweichung etwa gleich, während er in den TDG-/- Zellen leicht von 0.54% auf 0.64% zunimmt. Somit könnte tatsächlich etwas neu gebildetes $D_1^{-15}N_2$ -dC durch Deformylierung oder Decarboxylierung hinzugekommen sein.

Dies wird deutlicher in Abbildung 4.11 B und D. Hier erkennt man, dass sowohl im TDG+/- als auch im TDG-/- es zu einer leichten Zunahme an $D_2^{-15}N_2$ -dC nach 3 Tagen kommt.

Während der prozentuale Anteil an $D_1^{-15}N_2$ -dC in den TDG+/- Zellen nahezu gleich bleibt, bzw. um 0.02% leicht zunimmt, sieht man bei den TDG-/- Zellen eine deutliche Zunahme um 0.37%. Somit ist in Abwesenheit von TDG eine vermehrte Bildung an neuem $D_1^{-15}N_2$ -dC sichtbar. Die absoluten Mengen der Zunahme sind in Abbildung 4.12 dargestellt.

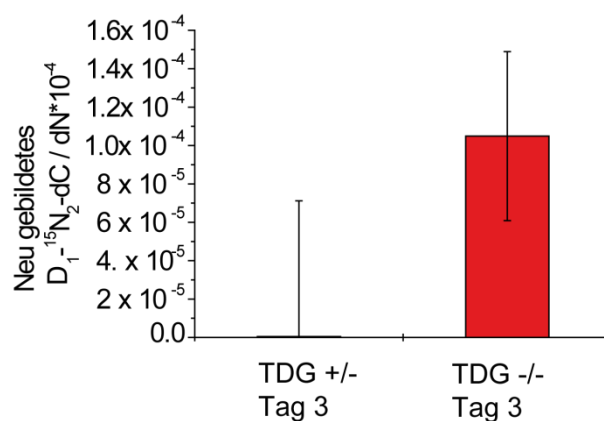


Abb. 4.12: Berechnung der absoluten Menge an neu gebildetem $D_1^{-15}N_2$ -dC im TDG -/- nach 3 Tagen.

In dieser Darstellungsweise wird deutlich, dass im TDG +/- nach 3 Tagen lediglich der chemische Hintergrund an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ detektierbar ist, der schon zum Zeitpunkt 0 Tage gemessen wurde. Im TDG -/- dagegen sieht man nach 3 Tage eine Zunahme an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ über dem chemischen Hintergrund. Bei dieser Zunahme von etwa $1.1 \cdot 10^{-4}$ pro dN (entspricht den 0.37% in Abb. 4.11) könnte es sich um durch Deformylierung oder Decarboxylierung neu gebildetes $D_1\text{-}^{15}\text{N}_2\text{-dC}$ handeln. Der absolute Wert liegt jedoch deutlich über den gemessenen Mengen an $D_1\text{-}^{15}\text{N}_2\text{-fdC}$, aus dem es durch Deformylierung entstanden sein müsste und erscheint somit auf dem ersten Blick etwas zu hoch. Allerdings wäre dies mit der Rolle von fdC als Intermediat in einem effizienten Deformylierungsmechanismus durchaus erklärbar. Dennoch sollte das Experiment wiederholt werden um sicher zu gehen, dass die gemessene Zunahme an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ tatsächlich derart hoch ist.

In diesem Experiment konnte durch die Anreicherung von $D_1\text{-}^{15}\text{N}_2\text{-dC}$ nach 3 Tagen Fütterung von $D_2\text{-}^{15}\text{N}_2\text{-dC}$ ein Hinweis auf eine Deformylierungs- oder Decarboxylierungsaktivität im TDG -/- der mESC gewonnen werden.

Parallel wurde nach einem geeigneten Modellsystem gesucht, in dem eine gesteigerte Decarboxylierungs- oder Deformylierungsaktivität beobachten werden könnte. Da die Verunreinigung an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ immernoch 3% betrug wäre ein System mit möglichst großer Decarboxylierungs- bzw. Deformylierungsaktivität wünschenswert. Durch eine stärkere Anreicherung von $D_1\text{-}^{15}\text{N}_2\text{-dC}$ wäre dann eine Zunahme über dem Hintergrund von 3% besser detektierbar. Ein System mit robust detektierbarem Signal würde zudem die Suche nach den beteiligten Faktoren erleichtern. Toni Pfaffeneder, der dieses Projekt begann unternahm erste Messungen in HeLa Zellen. Von Anfang an schienen Stammzellen der Maus jedoch das geeignetere System zu sein da hier viel höhere 5fdC und 5cadC Werte zu detektieren sind und somit auch mehr Deformylierung und Decarboxylierung zu erwarten ist. Im TDG -/- der mESC konnten wie weiter oben beschrieben erste Hinweise auf eine Deformylierungs- oder Decarboxylierungsaktivität gefunden werden.

Im Folgenden wurde nach Zellen gesucht, welche zu bestimmten Zeitpunkten einen Verlust an 5fdC oder 5cadC aufweisen, sodass zu diesem Zeitpunkt eine vermehrte Deformylierung oder Decarboxylierung beobachtet werden könnte.

Zuerst wurden *Epiblast like stem cells* (EpiLC) untersucht. Es handelt sich bei EpiLCs um pluripotente Zellen in einem geprimeten Zustand, welche durch eine Behandlung von

naïven murinen Stammzellen mit den beiden Wachstumsfaktoren FGF-2 und Activin A über 2-3 Tage generiert werden. Dies ahmt den Entwicklungsprozess nach, durch welchen naive preimplantatorische Epiblasten zu geprimten postimplantatorischen Epiblasten werden.^[165] Während dieser Zeit kommt es zur Ausbildung der epigenetischen Modifikationen. Um den besten Zeitpunkt für eine Verfütterung von D₂-¹⁵N₂-dC an die Epi LCs zu bestimmen, wurden die Zellen zu verschiedenen Zeitpunkten geerntet und hinsichtlich ihres Modifikationsgehaltes vermessen. (Abb. 4.13)

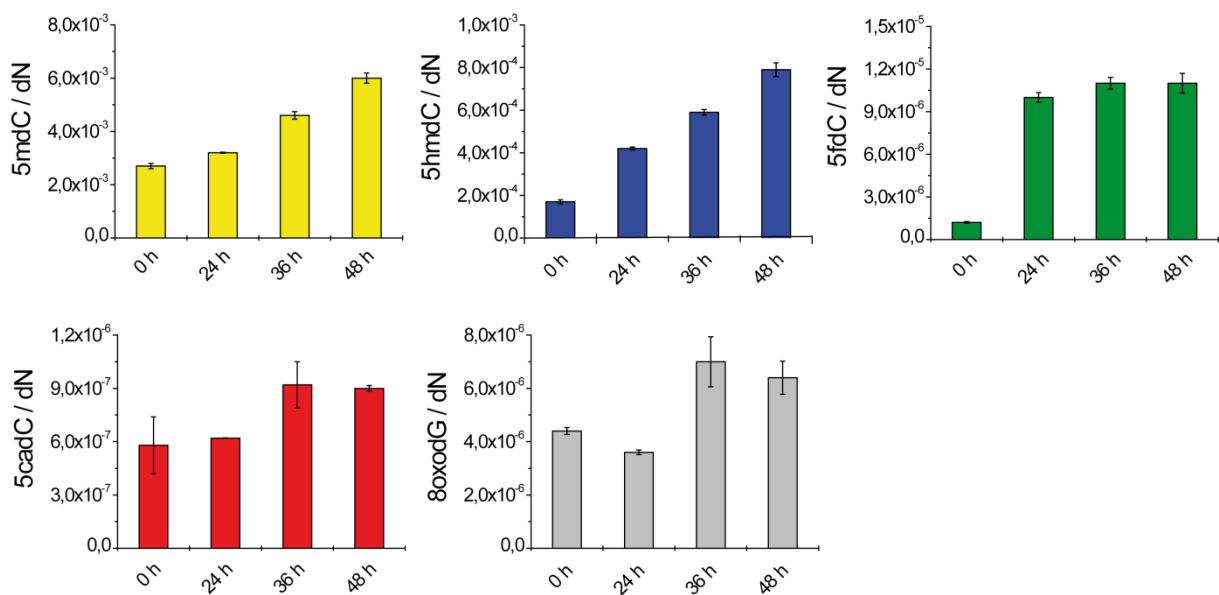


Abb. 4.13: Absolute Quantifizierung der Level an 5mdC, 5hmdC, 5fdC, 5 cadC und 8oxodG in Epi LCs während der ersten 48 h des Entwicklungsprozesses.

Man erkennt einen kontinuierlichen Anstieg der 5mdC sowie der 5hmdC Werte über die ersten 48 Stunden. Die 5fdC Werte steigen sprunghaft zwischen 0h und 24h an und bleiben anschließend etwa konstant. Die 5cadC Werte steigen erst mit 12 Stunden Verzögerung, nämlich zwischen 24 h und 36 h an und bleiben anschließend ebenfalls konstant. Unter Vernachlässigung des relativ großen Fehler für die 5cadC-Level zum Zeitpunkt 36 h sieht man einen leichten Abfall der 5cadC Werte zwischen 36 h und 48 h, dieser ist jedoch nicht signifikant. Dennoch scheint dies ein geeignetes System zu sein, um eine mögliche Decarboxylierungs- oder Deformylierungsaktivität zu beobachten. Die 5fdC und 5cadC Produktion ist nach 24 h bzw. 36h deutlich angekurbelt. Dies könnte einer vermehrten Demethylierung vorausgehen zumal zu diesem Zeitpunkt auch die Menge an 5mdC und damit die Häufigkeit fehlerhafter Methylierung zunimmt welche

anschließend wieder entfernt werden muss. Daher wurden zum Zeitpunkt 0 h und während der ganzen Kultivierung 100 μM $\text{D}_2\text{-}^{15}\text{N}_2\text{-dC}$ an die EpiLCs verfüttert und diese nach 48 h geerntet. Auch hier wurden die Mengen an $\text{D}_1\text{-}^{15}\text{N}_2\text{-mdC}$, $\text{D}_1\text{-}^{15}\text{N}_2\text{-hmdC}$ und $\text{D}_1\text{-}^{15}\text{N}_2\text{-fdC}$ ermittelt. (Abb. 4.14)

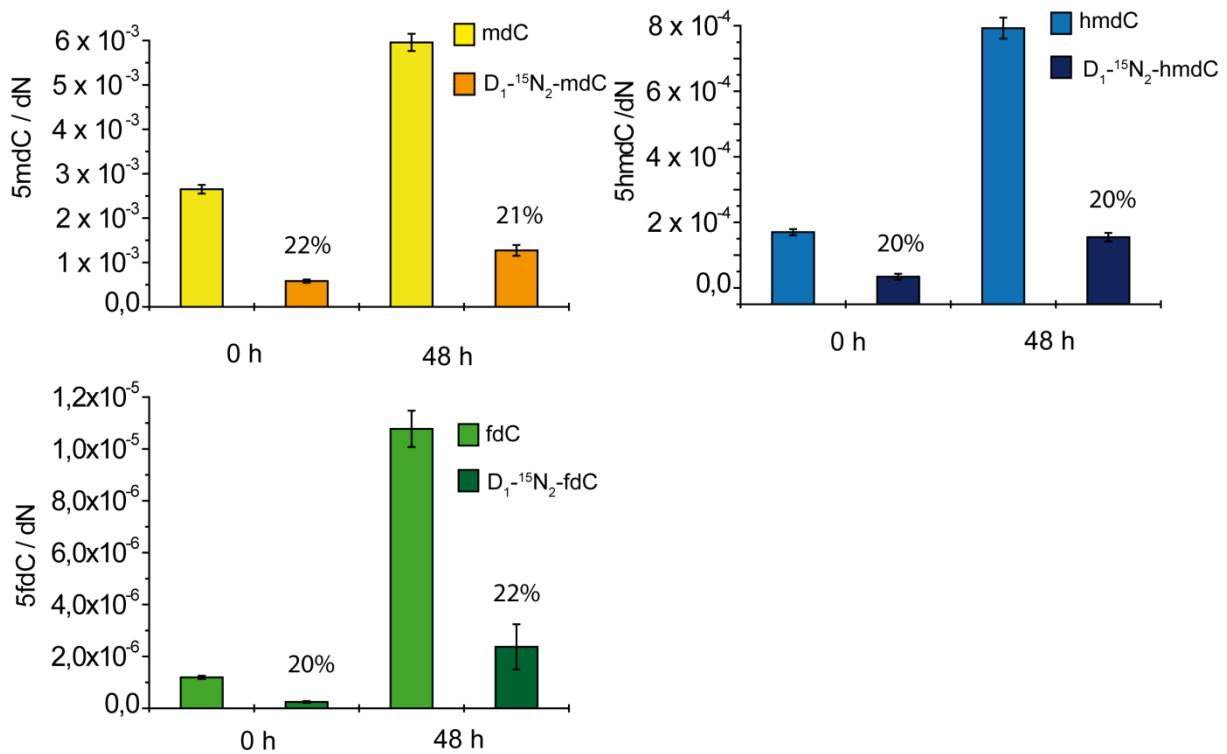


Abb. 4.14: Absolut quantifizierte Mengen an mdC, hmdC und fdC sowie berechnete Mengen der deuterierten Spezies $\text{D}_1\text{-}^{15}\text{N}_2\text{-mdC}$, $\text{D}_1\text{-}^{15}\text{N}_2\text{-hmdC}$ und $\text{D}_1\text{-}^{15}\text{N}_2\text{-fdC}$ jeweils zum Zeitpunkt 0 h und 48 h in EpiLCs.

Die Mengen an $\text{D}_1\text{-}^{15}\text{N}_2\text{-mdC}$, $\text{D}_1\text{-}^{15}\text{N}_2\text{-hmdC}$ und $\text{D}_1\text{-}^{15}\text{N}_2\text{-fdC}$ konnten allerdings nur relativ zu den Mengen an natürlichen Modifikationen errechnet werden. Eine absolute Quantifizierung ist auf Grund des Fehlens von geeigneten höher isotopenmarkierten Standards nicht möglich. Man erkennt auch hier, dass sich die deuterierten Modifikationen sehr ähnlichen zu den natürlichen Modifikationen verhalten. Im Vergleich zu den natürlichen Modifikationen finden sich immer zu etwa 20% die deuterierten Spezies. Bei genauerer Betrachtung erkennt man, dass die Menge an $\text{D}_1\text{-}^{15}\text{N}_2\text{-mdC}$ relativ zur Menge an natürlichem 5mdC nach der Ausdifferenzierung von 20% auf 21% leicht zunimmt bzw schwankt. Die Level an $\text{D}_1\text{-}^{15}\text{N}_2\text{-hmdC}$ scheinen sich konstant zu den 5hmdC Levels zu verhalten, während sich $\text{D}_1\text{-}^{15}\text{N}_2\text{-fdC}$ leicht zum Zeitpunkt 48h von 20% auf 22% angereichert hat. Unter Berücksichtigung der relativ

großen Standardabweichung ist dieser Trend jedoch nicht eindeutig.

Die Mengen an $D_2\text{-}^{15}\text{N}_2\text{-dC}$ und $D_1\text{-}^{15}\text{N}_2\text{-dC}$ sowie an dC konnten jeweils zum Zeitpunkt 0 h und zum Zeitpunkt 48 h absolut quantifiziert werden. (Abb. 4.15)

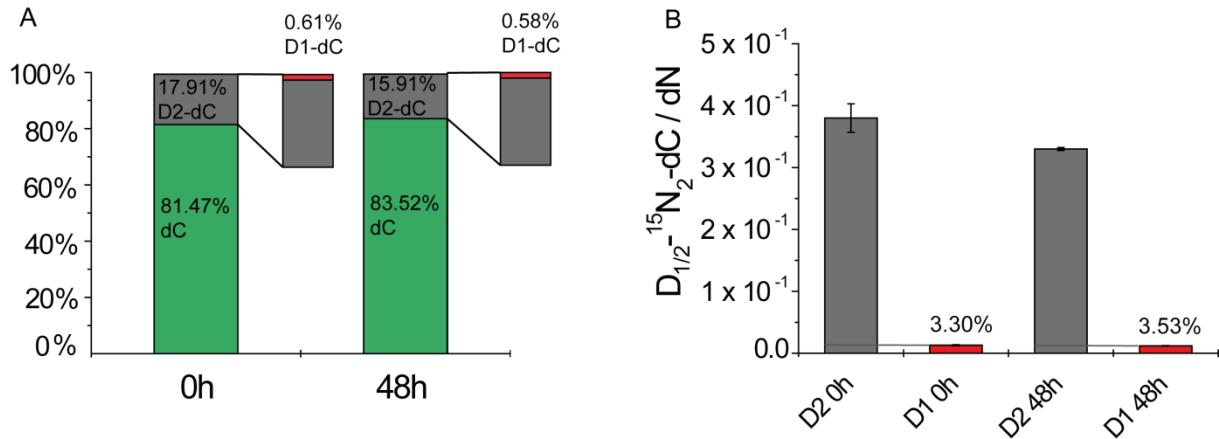


Abb.4.15: A) Relative Darstellung der Menge an dC, $D_2\text{-}^{15}\text{N}_2\text{-dC}$ und $D_1\text{-}^{15}\text{N}_2\text{-dC}$ in Prozent zum Zeitpunkt 0 h und 48 h. B) Absolute Quantifizierung der Mengen an $D_2\text{-}^{15}\text{N}_2\text{-dC}$ und $D_1\text{-}^{15}\text{N}_2\text{-dC}$ jeweils zum Zeitpunkt 0 h und 48 h. Berechnung der prozentualen Menge an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ im Vergleich zu $D_2\text{-}^{15}\text{N}_2\text{-dC}$ zum Zeitpunkt 0 h und nach 48 h.

In Abbildung 4.15 A sind die relativen Mengen an dC, $D_2\text{-}^{15}\text{N}_2\text{-dC}$ und $D_1\text{-}^{15}\text{N}_2\text{-dC}$ als prozentualer Anteil aller dC-Spezies (dC, $D_2\text{-}^{15}\text{N}_2\text{-dC}$, $D_1\text{-}^{15}\text{N}_2\text{-dC}$) dargestellt. In dieser Darstellungsweise erkennt man, dass der Anteil an dC nach 48 h etwas größer wird, während die Anteile an $D_2\text{-}^{15}\text{N}_2\text{-dC}$ und $D_1\text{-}^{15}\text{N}_2\text{-dC}$ etwas abnehmen. Der Anteil an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ nimmt dabei weniger stark ab als der Anteil an $D_2\text{-}^{15}\text{N}_2\text{-dC}$. Somit könnte tatsächlich etwas neu gebildetes $D_1\text{-}^{15}\text{N}_2\text{-dC}$ durch Deformylierung oder Decarboxylierung hinzugekommen sein.

Dies wird besser ersichtlich in Abbildung 4.15 B. Hier erkennt man, dass es zu einer leichten Abnahme an $D_2\text{-}^{15}\text{N}_2\text{-dC}$ nach 48 h kommt. Zum Zeitpunkt 0 h beträgt der prozentuale Anteil an $D_1\text{-}^{15}\text{N}_2\text{-dC}$ in $D_2\text{-}^{15}\text{N}_2\text{-dC}$ 3.40%. Dies kann als Hintergrund, zum größten Teil als chemischer Hintergrund aus der Synthese, bezeichnet werden. Nach 48 h kann nun allerdings ein prozentualer Anteil von 3.64% $D_1\text{-}^{15}\text{N}_2\text{-dC}$ in $D_2\text{-}^{15}\text{N}_2\text{-dC}$ ermittelt werden. Somit ist nach 48 h Fütterung, wenn man die 3.40% als Hintergrund setzt, zu 0.24% zusätzliches $D_1\text{-}^{15}\text{N}_2\text{-dC}$ entstanden. Diese 0.24% entsprechen einer absoluten Menge von 7.4×10^{-5} pro dN. Dies könnte theoretisch der Menge an durch Deformylierung oder Decarboxylierung neu entstandenem $D_1\text{-}^{15}\text{N}_2\text{-dC}$ entsprechen. Der absolute Wert liegt jedoch über der zu diesem Zeitpunkt gemessenen absoluten Menge an

$D_1\text{-}^{15}\text{N}_2\text{-fdC}$. Dieser beträgt $2.3 \cdot 10^{-6}$ pro dN und somit ist die Menge an neu entstandenem $D_1\text{-}^{15}\text{N}_2\text{-dC}$ etwas zu hoch, könnte jedoch wiederum durch einen vermehrten Umsatz des $D_1\text{-}^{15}\text{N}_2\text{-fdC}$ erklärt werden.

Zusammenfassend kann man sagen, dass man durch die Fütterung von $D_2\text{-}^{15}\text{N}_2\text{-dC}$ in EpiLCs ebenfalls ein Hinweis auf eine zusätzliche Bildung von $D_1\text{-}^{15}\text{N}_2\text{-dC}$ erhalten hat, ob dies allerdings tatsächlich durch Deformylierung oder Decarboxylierung entstanden ist, ist auf Grund der chemischen Verunreinigung der Ausgangsverbindung nicht ganz sicher zu belegen.

Im weiteren Verlauf des Projektes wurde in sogenannten *iNGNs* nach Deformylierungs- oder Decarboxylierungsaktivität gesucht. Es handelt sich dabei um humane induzierte pluripotente Stammzellen (hiPSCs) welche die beiden induzierbaren Transgene Neurogenin (NGN) 1 und 2 enthalten. Die Induktion dieser Transgene startet ein Neurogeneseprogramm wodurch sich die Zellen innerhalb von 4 Tagen simultan zu bipolaren Neuronen ausdifferenzieren.^[166] Besonders interessant ist in diesem Zusammenhang, dass die Proliferation mit der Differenzierung aufhört. Gebildetes mdC kann nicht mehr über passive Verdünnung entfernt werden. Damit sollten aktive Mechanismen bevorzugt ablaufen.

Um den idealen Zeitpunkt für die Fütterung von $D_2\text{-}^{15}\text{N}_2\text{-dC}$ zu finden wurden zunächst die globalen Level der epigenetischen Modifikationen 5mdC, 5hmdC, 5fdC und 5cadC während der Ausdifferenzierung zu Neuronen bestimmt. (Abb. 4.16)

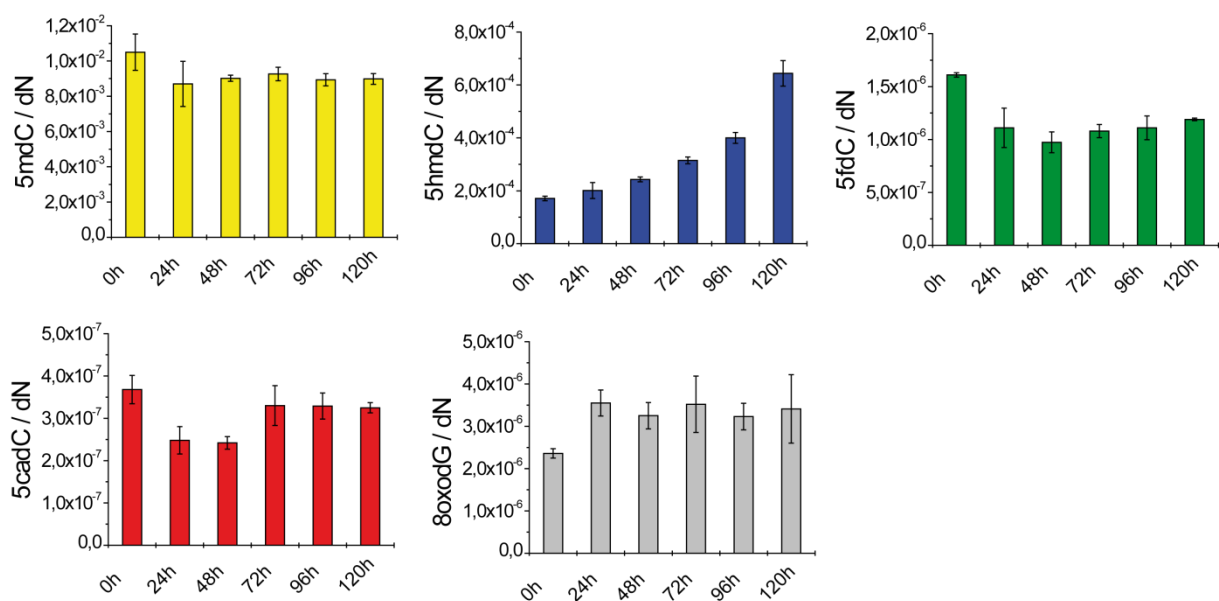


Abb. 4.16: Absolute Quantifizierung der Level an 5mdC, 5hmdC, 5fdC, 5 cadC und 8oxodG während der Ausdifferenzierung von *iNGNs* zu Neuronen.

Während hier die mC Werte annähernd konstant sind, oder zumindest zwischen 0 h und 24 h nur leicht aber nicht signifikant abnehmen, nehmen die 5hmdC Werte kontinuierlich über die Zeit zu. Die 5fdC und 5cadC Level zeigen eine deutliche und signifikante Abnahme zwischen 0h und 24 h. Erst zwischen 48 h und 72 h kommt es erneut zu einer Zunahme von 5fdC und 5cadC. Dies ist das erste Mal, dass eine deutliche Abnahme der globalen 5fdC und 5cadC Level während der Ausdifferenzierung von Zellen detektiert werden konnte. Somit findet in diesen Zellen möglicherweise zwischen dem Zeitpunkt 0 h und 24 h bzw. 48 h neben aktiver Demethylierung über Basenexzisionsreparatur auch vermehrt Deformylierung oder Decarboxylierung statt. Dies macht die Ausdifferenzierung dieser Zellen zu einem geeigneten System, um eine Anreicherung von $D_1\text{-}^{15}\text{N}_2\text{-dC}$ beobachten zu können. Idealerweise würde in diesen Zellen ein TDG *Knockout* vorgenommen werden. Dadurch würde der Anteil an aktiver Demethylierung über Deformylierung und Decarboxylierung durch das Fehlen von TDG und damit der Basenexzisionsreparatur noch verstärkt werden.

Jeweils zu Beginn der Kultivierung wurde 100 μM $D_2\text{-}^{15}\text{N}_2\text{-dC}$ an die *iNGNs* verfüttert. Zu diesem Zeitpunkt befanden sich die Zellen noch in einem pluripotenten Status. Anschließend wurde durch die Induktion der Transgene Neurogenin (NGN) 1 und 2 die Ausdifferenzierung gestartet. Zum diesem Zeitpunkt wurde die Fütterung von $D_2\text{-}^{15}\text{N}_2\text{-dC}$ gestoppt. Die Zellen wurden 2 Tage ausdifferenziert und zum Zeitpunkt 0 h und 1 Tag bzw. 2 Tage der Ausdifferenzierung geerntet.

In einem ersten Experiment wurde das verfütterte $D_2\text{-}^{15}\text{N}_2\text{-dC}$ nur zu 2.5% in die DNA der *iNGNs* eingebaut. Wegen der geringen Inkorporierung wird die Quantifizierung von $D_1\text{-}^{15}\text{N}_2\text{-dC}$ zusätzlich erschwert. Neben dem chemischen Hintergrund an dC+3 Spezies, welcher aus der Synthese des $D_2\text{-}^{15}\text{N}_2\text{-dC}$ kommt, spielt bei so geringen Einbauraten auch der Hintergrund aus an natürlichem dC+3 aus der DNA eine Rolle. Auf Grund der sehr geringen Gesamtmenge an eingebautem $D_1\text{-}^{15}\text{N}_2\text{-dC}$ in der DNA liegt dieser hier kaum über dem Beitrag an dC+3 aus der natürlichen DNA, welche zu 0.106% dC+3 enthält. Hier steht man also vor einer zusätzlichen Herausforderung. Da das durch Decarboxylierung oder Deformylierung neu gebildete $D_1\text{-}^{15}\text{N}_2\text{-dC}$ in einer wesentlich kleineren Größenordnung erwartet werden würde, wird mit diesen Einbauraten eine verlässliche Quantifizierung unmöglich. Auch eine Quantifizierung von $D_1\text{-}^{15}\text{N}_2\text{-mdC}$, $D_1\text{-}^{15}\text{N}_2\text{-hmdC}$ und $D_1\text{-}^{15}\text{N}_2\text{-fdC}$ war bei 2.5% Inkorporierung nicht möglich.

Durch Fütterung der *iNGNs* mit $D_2\text{-}^{15}\text{N}_2\text{-dC}$ über einen längeren Zeitraum vor der

Ausdifferenzierung konnte eine Einbaurrate von knapp über 7% erzielt werden. Bei einem Einbau von 7% D_2 - $^{15}N_2$ -dC konnten nun wie zuvor in den EpiLCs die Mengen an D_1 - $^{15}N_2$ -mdC, D_1 - $^{15}N_2$ -hmdC und D_1 - $^{15}N_2$ -fdC ermittelt werden. (Abb. 4.17)

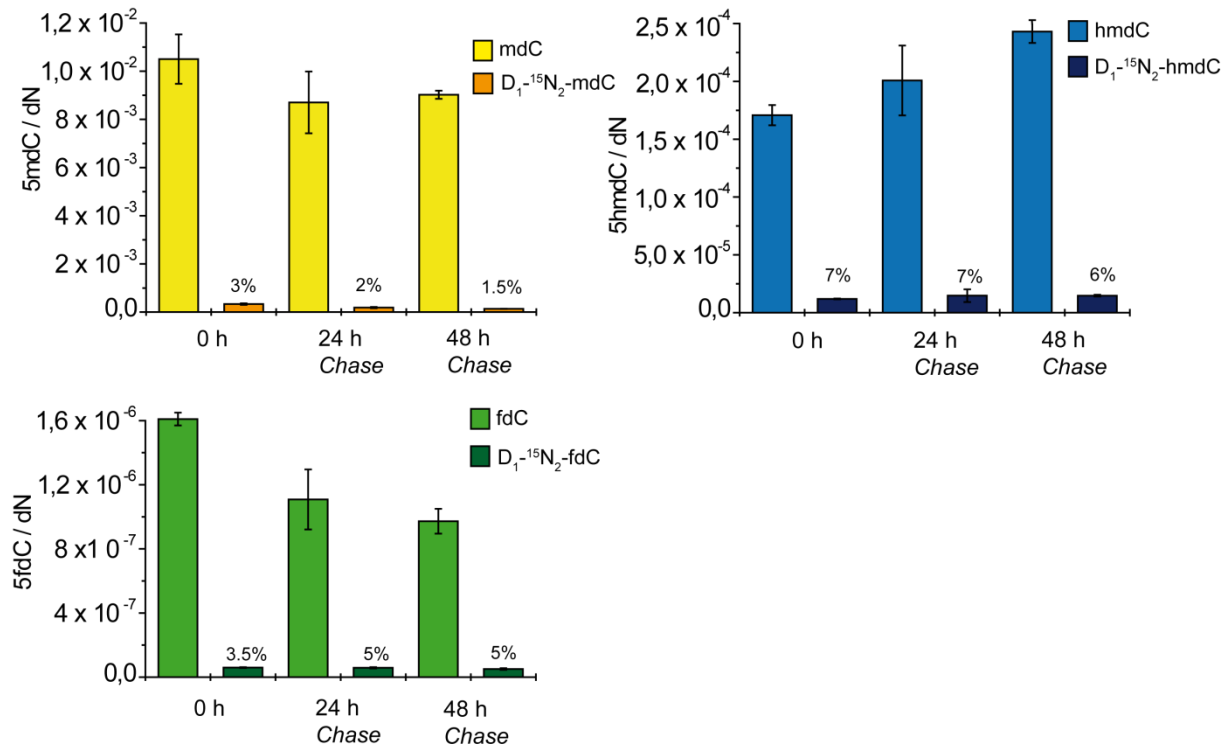


Abb. 4.17: Absolut quantifizierte Mengen an mdC, hmdC und fdC sowie berechnete Mengen der deuterierten Spezies D_1 - $^{15}N_2$ -mdC, D_1 - $^{15}N_2$ -hmdC und D_1 - $^{15}N_2$ -fdC jeweils zum Zeitpunkt 0 h, 24 h und 48 h nach Beendigung der Verfütterung (=Chase) von D_2 - $^{15}N_2$ -dC in iNGNs.

Auch hier konnte keine absolute Quantifizierung der Mengen an D_1 - $^{15}N_2$ -mdC, D_1 - $^{15}N_2$ -hmdC und D_1 - $^{15}N_2$ -fdC erfolgen. Die Mengen sind relativ zu den gemessenen Mengen an 5mdC, 5hmdC und 5fdC berechnet worden. Man erkennt, dass die Menge an D_1 - $^{15}N_2$ -mdC relativ zur Menge an natürlichem 5mdC nach Beendigung der Verfütterung und zu Beginn der Ausdifferenzierung von 3% auf 1.5% leicht abnimmt. Die D_1 - $^{15}N_2$ -hmdC Level scheinen sich konstant zu den 5hmdC Levels zu verhalten, während sich D_1 - $^{15}N_2$ -fdC leicht zum Zeitpunkt 48 h nach Ende der Verfütterung von D_2 - $^{15}N_2$ -dC von 3.5% auf 5% angereichert hat. Diese Anreicherung ist auch unter Berücksichtigung der Fehler signifikant. Wie schon zuvor bei den TDG-/- Zellen scheinen markierte D_1 - $^{15}N_2$ -hmdCs nach 24 h bzw 48 h etwas bevorzugt oxidiert zu werden. Dies könnte bedeuten, dass es sich hier um deformyliertes oder decarboxyliertes Substrat handelt welches nach einer

Remethylierung bevorzugt oxidiert und anschließend wieder entfernt werden soll.

Auch bei einer Inkorporationsrate von 7% liegen die Mengen an detektierbarem $D_1\text{-}^{15}N_2\text{-dC}$ jedoch nicht deutlich über dem biologischen Hintergrund aus der DNA, sodass auch hier eine Quantifizierung von $D_1\text{-}^{15}N_2\text{-dC}$ ein falsches Ergebnis liefern würde. Möglicherweise könnte durch eine Steigerung der verfütterten Stoffmengenkonzentration von 100 μM auf 200 μM $D_2\text{-}^{15}N_2\text{-dC}$ der Einbau in die DNA verbessert werden.

Eine weitere Verbesserungsmöglichkeit könnte in der Verwendung mehrerer oder anderer Isotopen zur Markierung der Ursprungsverbindung liegen. Würde an Stelle eines Massenunterschieds von 4 u ($dC+4$) durch zweimal ^{15}N -Stickstoff und zweimal Deuterium, ein $dC+5$ durch dreimal ^{15}N -Stickstoff oder besser noch dreimal ^{13}C -Kohlenstoff und zweimal Deuterium erreicht werden, würde der Hintergrund, welcher durch die natürlichen Nukleoside verursacht wird, verschwindend gering werden. Nach einer Decarboxylierung oder Deformylierung würde dann $dC+4$ erhalten werden. Bei $dC+4$ macht der natürliche Anteil nur noch etwa 0.007% aller dCs aus, bei $dC+5$ sind es sogar nur noch 0.0003%. Somit würde der Wert deutlich unter dem erwarteten Wert für $D_1\text{-}^{15}N_2\text{-dC}$ liegen und der biologische Hintergrund würde die Messungen selbst bei geringen Einbauraten nicht mehr verfälschen.

Zudem könnte mit einer alternativen Syntheseroute und dem Umgehen der Verwendung von ^{15}N -Harnstoff, welcher kommerziell nur mit einer Isotopenanreicherung von 98% erhältlich ist, eine wesentlich höhere Isotopenanreicherung des $dC+4$ erreicht werden. Dadurch könnte ein chemischer Hintergrund an $dC+3$ von deutlich weniger als 3% erreicht werden.

Eine weitere Möglichkeit die detektierte Menge an $D_1\text{-}^{15}N_2\text{-dC}$ zu erhöhen, wäre wie schon erwähnt in den EpiLC oder *iNGN*-Zellen TDG auszuschalten. Dadurch könnte wie weiter oben beschrieben in den Stammzellen der Maus, der Effekt verstärkt und anschließend die Änderung unter Abzug des Hintergrundes berechnet werden.

Parallel zur Fütterung des $D_2\text{-}^{15}N_2\text{-dC}$ wurden Fütterungsstudien von fluorierten Nukleosiden vorgenommen. Im Gegenteil zu $D_2\text{-}^{15}N_2\text{-dC}$ sind diese nicht natürlich. An Stelle der beiden Deuteriumatome im $D_2\text{-}^{15}N_2\text{-dC}$, an Positionen die zuvor mit Wasserstoffatomen besetzt waren, findet sich ein sterisch zwar sehr ähnliches, elektronisch aber zu Wasserstoff sehr unterschiedlicheres Fluoratom an der 2'-Position

des Zuckers. Dies bietet allerdings zwei wesentliche Vorteile. Zum einen wird durch das Fluoratom der Übergangszustand während der Deglykosylierung so sehr destabilisiert, dass in die DNA eingebaute fluorierte Nukleoside nicht mehr durch TDG und BER entfernt werden. Zum anderen sind die fluorierten Nukleoside massenspektrometrisch durch den hohen Massenunterschied sehr gut von den natürlichen Nukleosiden unterscheidbar. Zudem hat man hier nicht das Problem des chemischen Hintergrundes aus der Synthese.

Mit den fluorierten Substanzen konnte von Katharina Iwan und Arne Schröder ebenfalls deutliche Hinweise auf eine Deformylierungsaktivität in mESC gefunden werden.

4.2.3 Zusammenfassung und Projektbeiträge

Die Synthese des $D_2\text{-}^{15}N_2\text{-}2'$ -Desoxycytidines wurde erstmals von Toni Pfaffeneder und Arne Schröder durchgeführt. René Rahimoff optimierte die Synthese durch einen nachfolgenden Deuterium-Anreicherungsschritt. Die Entwicklung der Quantifizierungsmethode basierte auf Vorarbeiten von Toni Pfaffeneder und wurde von mir weiter für die Anwendung optimiert. Die Zellkulturversuche wurden von Fabio Spada durchgeführt, ebenso die Isolation der genomischen DNA. Der enzymatische Verdau sowie die Messung, Auswertung und Interpretation der Daten wurden von mir vorgenommen. Die Daten zum TDG *Knockout* zu verschiedenen Zeitpunkten wurden von René Rahimoff generiert, das Projekt wurde von ihm übernommen.

4.2.4 Material und Methoden

DNA Isolation

Die Isolation der DNA erfolgte nach dem unter 3.3 beschriebenen Protokoll.

Verdau

Der Verdau der genomischen DNA erfolgte ebenfalls nach dem unter 3.3 beschriebenen Protokoll.

LC-MS/MS-Methode

Die quantitative LC-ESI-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte wie unter 4.1.5 beschrieben mit einem 1290 UHPLC-System von *Agilent*, welches über einen UV Detektor verfügt und mit einem 6490 Triple-Quadrupol Massenspektrometer von *Agilent* verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8-Säule (*Agilent*, 2.7 μm , 2.1 mm \times 150 mm) bei 30 °C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0.0085% (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0.35 mL/min: 0 \rightarrow 5 min; 0 \rightarrow 3.5% (v/v) MeCN; 5 \rightarrow 6.9 min; 3.5 \rightarrow 5% MeCN; 6.9 \rightarrow 7.2 min; 5 \rightarrow 80% MeCN; 7.2 \rightarrow 10.5 min; 80% MeCN; 10.5 \rightarrow 11.3 min; 80 \rightarrow 0% MeCN; 11.3 \rightarrow 13 min; 0% MeCN.

Die Quellparameter waren dieselben wie unter 4.1.5. Die für die jeweiligen Analyten spezifischen Parameter sind in Tabelle 4.2 zusammengefasst.

Tabelle 4.2: Substanz abhängige LC/MS-MS Parameter zur Quantifizierung der markierten Nukleoside in genomischer DNA

Verbindung	Precursor Ion (m/z)	MS1 Auflösung	Produkt Ion (m/z)	MS2 Auflösung	Dwell time [ms]	CE (V)	CAV (V)	Polarität
dC	228.1	Unit	112.05	enhanced	20	1	0	positiv
[d ₂ - ¹⁵ N ₂]-dC	232.1	Unit	115.1	Unit	20	30	5	positiv
[d ₁ - ¹⁵ N ₂]-dC	231.1	Unit	116.1	Unit	20	30	5	positiv
[¹⁵ N ₃ - ¹³ C ₉]-dC	240.1	Unit	119.1	Unit	20	30	5	positiv
[d ₁ - ¹⁵ N ₂]-mdC	245.1	Unit	129.1	Unit	35	40	5	positiv
[d ₃ - ¹⁵ N ₂]-mdC	247.1	Unit	131.1	Unit	35	40	5	positiv
mdC	242.1	Unit	126.1	Unit	35	40	5	positiv
[d ₁ - ¹⁵ N ₂]-hmdC	261.1	Unit	145.1	Unit	40	27	5	positiv
hmdC	258.1	Unit	142.06	Unit	40	27	5	positiv
[d ₁ - ¹⁵ N ₂]-fdC	259.1	Wide	143.1	Wide	40	5	5	positiv
fdC	256.1	Wide	140.1	Wide	40	5	5	positiv
[d ₁ - ¹⁵ N ₂]-cadC	275.1	Wide	159.04	Wide	40	5	5	positiv
cadC	272.1	Wide	156.04	Wide	40	5	5	positiv
[¹⁵ N ₅]-8-oxo-dG	289.08	Wide	173.04	Wide	20	9	7	positiv
8-oxo-dG	284.1	Wide	168.05	Wide	20	9	7	positiv

Validierung der LC-MS Methode

Die Validierung erfolgte wie unter 4.1.5 beschrieben.

Auswertung

Die Auswertung erfolgte wie unter 3.3 beschrieben mit dem Programm Mass Hunter und Excel, die Darstellung der Daten erfolgte mit Origin und Adobe Illustrator.

4.3 Desaminierung von Methylcytosin zu Thymin

4.3.1 Prolog

In Kapitel 4.2 wurde die Möglichkeit einer aktiven DNA Demethylierung über Deformylierung von 5fdC oder Decarboxylierung von 5cadC diskutiert. Diese Option wurde verglichen mit einer aktiven DNA Demethylierung über ein Ausschneiden von 5fdC und 5cadC mit einer Glykosylase und anschließender Basenexzisionsreparatur. Tatsächlich gibt es aber auch noch eine weitere denkbare Möglichkeit DNA aktiv zu demethylieren.

Genomisches 5mdC und 5hmdC könnten auch durch Cytosin Desaminasen der AID/APOBEC Familie oder durch die *de novo* DNA Methyltransferasen Dnmt 3a/b zu dT und 5hmU desaminiert werden. [9, 35, 36, 92, 101-103, 167] Die resultierenden T:G und hmU:G Fehlpaarungen würden anschließend über Basenexzisionsreparatur behoben werden könnten. Weiterhin könnten durch eine Desaminierung von dC innerhalb von 5mdC enthalten Sequenzen durch die AID Deaminase U:G Fehlpaarungen generiert werden, welche dann über *long patch* BER oder nicht kanonische *mismatch* Reparatur (ncMMR) repariert werden würden. Dies würde zu einer indirekten Reparatur der umgebenden 5mdCs führen.[32, 168] Diese Möglichkeiten sind allerdings nicht bewiesen und werden kontrovers diskutiert[169-172] zumal der Mechanismus über hoch mutagene Zwischenstufen verlaufen würde.[173]

Mit Hilfe der Fütterung von ^{13}C und D_3 markiertem Methionin ($\text{D}_3\text{-}^{13}\text{C-Met}$), welches durch Metabolisierung zu $\text{D}_3\text{-}^{13}\text{C-SAM}$ wird und schließlich zu einer Markierung der Methylgruppe führt, kann mit Hilfe von Isotopenverfolgungsexperimenten die Menge oxidierten Produkten bestimmt werden (Abb. 4.18). [165]

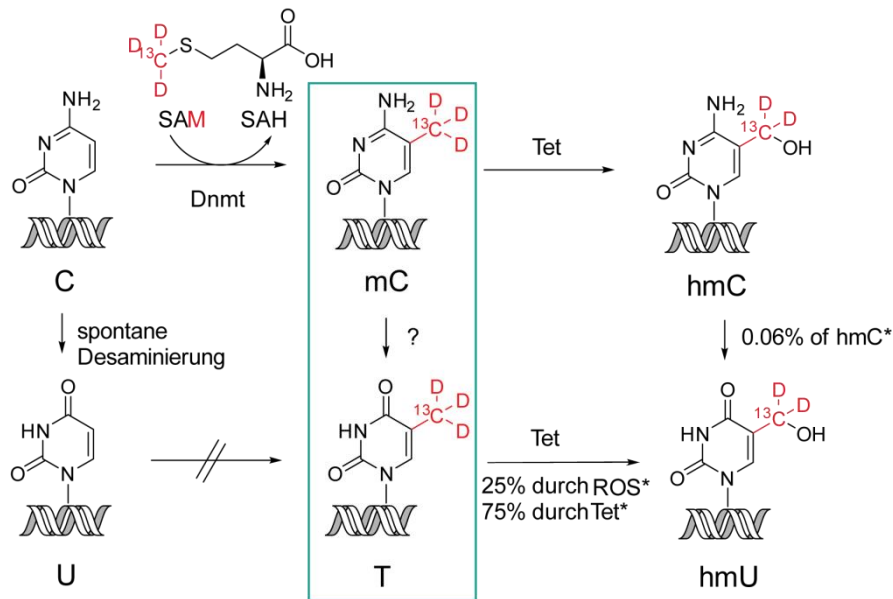


Abb. 4.18: Schematische Darstellung der Inkorporierung der Isotopenmarkierung während der Methylierung mit Hilfe von $D_3-^{13}C$ -SAM und anschließende Verfolgung der Markierung zur Bestimmung des Anteils an genomischer Desaminierung von 5mdC zu dT. (*) kennzeichnet Ergebnisse aus früheren Experimenten.^[165]

Zudem gibt die Menge an gebildetem $D_3-^{13}C$ -dT und $D_2-^{13}C$ -5hmdU Aufschluss über den Anteil der Desaminierung bei der Entstehung der isotope markierten Nukleoside in genomischer DNA. So scheint 5hmdU wie unter 3.3 gezeigt zum größten Teil über die Oxidation von dT durch die Tet Enzyme zu entstehen. Zu einem geringeren Anteil entsteht es aber auch aus ROS. Lediglich ein sehr kleiner Teil konnte durch genomische Desaminierung von 5hmdC gebildet werden.^[165] In welchem Ausmaß es zu einer genomischen Desaminierung von 5mdC zu dT kommt, konnte bisher nicht geklärt werden. Diese Frage wird im Folgenden beleuchtet. Eine Bildung von $D_3-^{13}C$ -dT über eine spontane Desaminierung von dC zu dU kann auf Grund der fehlenden Isotopenmarkierung hier ausgeschlossen werden.

4.3.2 Ergebnisse und Diskussion

Basierend auf der unter 3.3. beschriebenen und veröffentlichten Quantifizierungsmethode zur Bestimmung von schwerisotopen-markierten Produkten der D_3 - ^{13}C -S-Methionin (D_3 - ^{13}C -Met)-Fütterung [165] wurde zusätzlich eine Methode zur sensitiven Quantifizierung von D_3 - ^{13}C -dT entwickelt.

Eine besondere Herausforderung bei der Quantifizierung von markiertem dT besteht darin, dass auf Grund von sehr großen Mengen an natürlich vorliegendem dT in der DNA der Hintergrund und damit die Ionensuppression sehr hoch sind. So finden sich in der DNA zu 0.01 % Spezies mit einem Massenunterschied von 4 Einheiten zu dT (dT+4). Diese entsprechen somit in ihrer Masse dem aus der Fütterung entstandenen D_3 - ^{13}C -dT und müssen daher voneinander unterschieden werden. Die Ionensuppression wird durch die relativ späte Elution von dT bei der Analyse auf der UHPLC durch die Elution vieler anderer Analyten zu einer ähnlichen Retentionszeit noch verstärkt. Bei einer späten Elution ist außerdem der Gehalt an Acetonitril im Elutionspuffer schon relativ hoch, wodurch auch Verunreinigungen besser ionisiert werden, was den Hintergrund zusätzlich erhöht. Zudem sollte die Quantifizierung des D_3 - ^{13}C -dT mit Hilfe des höheren Isotopologes $^{13}C_{10}$ - $^{15}N_2$ -dT als Standard erfolgen. Auch dieses höhere Isotopolog trug durch Verunreinigung an dT+4 zu einem geringeren Teil zur Ionensuppression bei.

Bei ersten Versuchen D_3 - ^{13}C -dT aus genomischer DNA zu quantifizieren wurden daher sehr große Mengen an DNA verdaut. Es konnte herausgefunden werden, dass auch Komponenten des Total-Verdaus wie Enzyme oder Inhibitoren zur Ionensuppression beitragen. Durch Verwendung geringerer Mengen an Verdauenzymen konnten bessere Signale erzielt werden. Um eine Desaminierung nach der DNA-Isolation zu verhindern wurde beim Verdau der Cytidin-Deaminase Inhibitor Tetrahydrouridin (THU) zugesetzt. Durch Reduktion der Menge an verwendetem THU konnte auch hier das Signal für D_3 - ^{13}C -dT verbessert werden. Durch eine Optimierung der Verdaubedingungen konnte schließlich mit einer wesentlich geringeren Menge an DNA das Signal zu Rausch Verhältnis deutlich verbessert und ein solides Signal für D_3 - ^{13}C -dT erhalten werden, welches eine sensitive Quantifizierung ermöglichte.

Bei ersten Messungen mit der optimierten Methode fiel zunächst eine unerwartet hohe Menge an $D_3\text{-}^{13}\text{C}$ -dT in Wildtyp von Stammzellen der Maus (mESC) auf. Verglichen mit den absoluten Mengen an 5hmdU im Genom schienen die Werte für $D_3\text{-}^{13}\text{C}$ -dT viel zu hoch zu sein. Dies warf die Frage auf, ob die große Menge andetektiertem $D_3\text{-}^{13}\text{C}$ -dT tatsächlich von einer Desaminierungsreaktion in genomischer DNA stammt, über eine Desaminierung im löslichen Nukleosidpool und Reinkorporierung oder auf anderen Wegen in die DNA gelangt. In Abbildung 4.19 sind mögliche Wege der genomischen Desaminierung von $D_3\text{-}^{13}\text{C}$ -5mdC sowie einer Desaminierung im löslichen Pool der Nukleoside mit anschließender Reinkorporierung dargestellt.

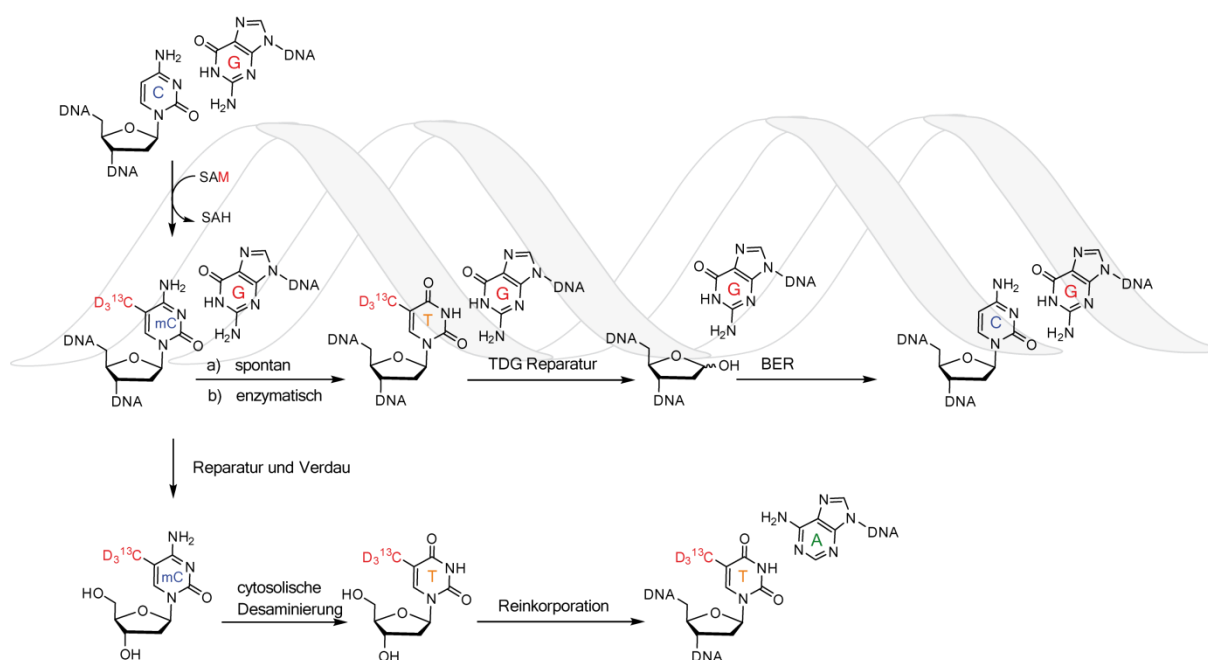


Abb. 4.19: Darstellung einer genomischen Desaminierung von $D_3\text{-}^{13}\text{C}$ -5mdC zu $D_3\text{-}^{13}\text{C}$ -dT sowie einer Desaminierung im löslichen Nukleosidpool mit anschließender Reinkorporierung in die DNA. Während eine genomische Desaminierung nach Basenexzisionsreparatur zu einer C:G Paarung führt, würde der Einbau von dT nach einer Desaminierung im Cytosol zu einem T:A Basenpaar führen.

Neben einer Desaminierung von 5mdC zu dT in der DNA könnte 5mdC auch durch Reparatur und Verdau als Nukleosid in den löslichen Nukleosidpool gelangen und dort durch Cytidin-Desaminasen desaminiert und im Anschluss re-inkorporiert werden.

Um zu überprüfen ob die Bildung von $D_3\text{-}^{13}\text{C}$ -dT tatsächlich direkt von 5mdC bzw. von SAM abhängig ist, wurden die Mengen an $D_3\text{-}^{13}\text{C}$ -dT in einer Dnmt Triple *Knockout* (Dnmt TKO) Zelllinie bestimmt. In dieser Zelllinie fehlen im Vergleich zum Dnmt

Wildtyp (Dnmt wt) Zellen die DNA-Methyltransferasen Dnmt 1, Dnmt 3a und Dnmt 3b. Dadurch sind die Dnmt TKO Zellen nicht mehr in der Lage 5-Methylcytosin zu bilden. Sofern $D_3\text{-}^{13}\text{C-dT}$ über $D_3\text{-}^{13}\text{C-5mdC}$ gebildet wird, sollte im Dnmt TKO also auch kein $D_3\text{-}^{13}\text{C-dT}$ mehr detektierbar sein. (Abb. 4.20)

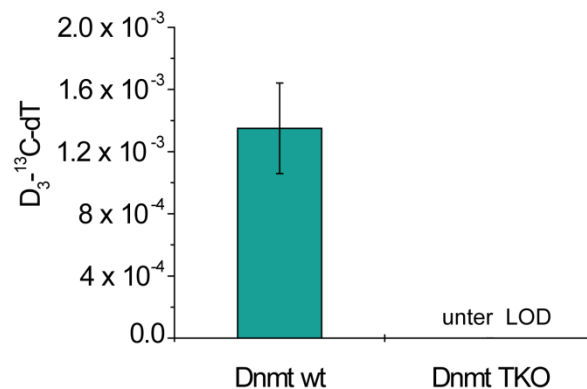


Abb. 4.20: Absolute Quantifizierung der Mengen an $D_3\text{-}^{13}\text{C-dT}$ in Dnmt wt Zellen und Dnmt TKO Zellen. Im Dnmt TKO lagen die Mengen unter dem Detektionslimit (LOD), N=3.

Dies ist im Dnmt TKO tatsächlich der Fall. Es konnte kein $D_3\text{-}^{13}\text{C-dT}$ mehr detektiert werden, die Werte lagen unter dem Detektionslimit (LOD). Somit kann ausgeschlossen werden, dass noch weitere Dnmt- und SAM-unabhängige Pfade zur relativ hohen Menge an $D_3\text{-}^{13}\text{C-dT}$ beitragen. Die Bestimmung des $D_3\text{-}^{13}\text{C-dT}$ -Level im Dnmt TKO klärt allerdings nicht die Frage nach genomischer Desaminierung oder Desaminierung im löslichen Nukleosidpool, da durch einen Dnmt TKO beide Wege unterbrochen sind. Hier sollte allerdings erwähnt werden, dass die Menge an $D_3\text{-}^{13}\text{C-dT}$ in den Dnmt wt Zellen stark variiert. Dies ist abhängig vom jeweiligen Methylierungsgrad der Zelle. Daher wurden im Folgenden die $D_3\text{-}^{13}\text{C-dT}$ Werte, sofern zu vergleichende Zelllinien oder Zeitreihen Unterschiede im Methylierungsgrad aufwiesen, auf 5mdC referenziert. Finden sich keine Unterschiede im Methylierungsgrad wurde auf eine Referenzierung verzichtet.

Zur Klärung der Frage, ob es sich bei der großen Menge an $D_3\text{-}^{13}\text{C-dT}$ um genomisch oder cytosolisch- desaminiertes $D_3\text{-}^{13}\text{C-mdC}$ handelt, wurde eine Titrierung von dT vorgenommen. Mit Hilfe immer größerer Mengen an unmarkiertem Desoxythymidin, welches als Konkurrenz im löslichen Pool an Nukleosiden fungiert, sollte mit

zunehmender Menge an dT immer weniger markiertes $D_3\text{-}^{13}\text{C-dT}$ eingebaut werden. (Abb. 4.21)

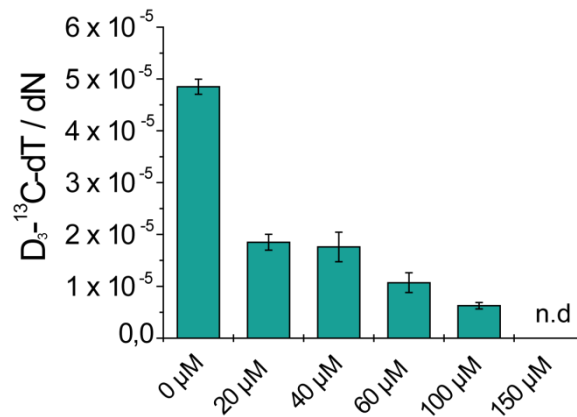


Abb. 4.21: Absolute Quantifizierung von $D_3\text{-}^{13}\text{C-dT}$ mit zunehmender Menge an natürlichen dT als Konkurrenzsubstrat in löslichen Nukleosidpool beim Einbau in die DNA.

Schon bei einer Zugabe von 20 μM dT kann nur noch 2/5 der ursprünglichen Menge an $D_3\text{-}^{13}\text{C-dT}$ detektiert werden. Bei einer Menge von 60 μM dT ist noch 1/5 der Ausgangsmenge sichtbar. Ab 150 μM kann zwar noch ein Signal für $D_3\text{-}^{13}\text{C-dT}$ detektiert werden, eine Quantifizierung dieser sehr geringen Mengen ist allerdings nicht mehr möglich, da hier die Quantifizierungsgrenze unterschritten wird. Somit kann nicht mit Sicherheit angenommen werden, dass selbst bei einer Konkurrenz mit 150 μM an natürlichem Thymidin ein geringer Rest an $D_3\text{-}^{13}\text{C-dT}$ bestehen bleibt, welcher der Anteil aus genomischer DNA sein könnte. Damit eignet sich eine Menge von 100 μM dT am besten für Konkurrenzexperimente. Hier ist zwar noch nicht die volle Absättigung erreicht, aber man sieht einen deutlichen Effekt.

Um weitere Aufschlüsse über den Anteil an genomischer Desaminierung zu bekommen wurde eine *Doppelknockout*-Zelllinie der beiden cytosolischen Desaminasen Desoxycytidylat-Desaminase (Dctd) und Cytosin-Desaminase (Cda) (Abb. 4.22) unter Verwendung des CRISPR/Cas9 Systems erzeugt.

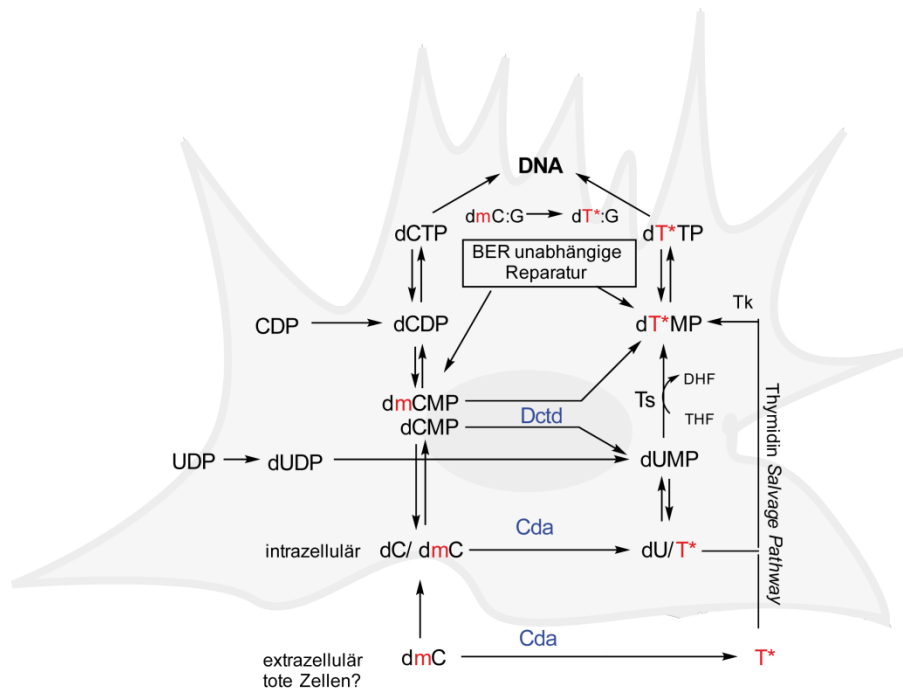


Abb. 4.22: Desaminierungswege der Cytosin Desaminase (Cda) und Desoxycytidylat Desaminase (Dctd), sowie der Thymidin Kinase (Tk) und der Thymidylat Synthase (Ts). Durch Ausschalten der cytosolischen Desaminasen Cda und Dctd sollte die cytosolische Desaminierung zum Erliegen kommen. Mit rot gekennzeichnet ist jeweils die Isotopenmarkierung.

In Anwesenheit der Desaminasen Desoxycytidylat Desaminase (Dctd) und Cytosin Desaminase (Cda) sollten die in der Abbildung 4.22 dargestellten Wege der Desaminierung von 5mdC/5mdCMP im Nukleosidpool geblockt sein und dadurch der Anteil an genomischer Desaminierung detektierbar sein. Nach der Validierung des Doppelknockouts der beiden Desaminasen (CD DKO) wurde die Menge an $D_3\text{-}^{13}\text{C-dT}$ in den unbehandelten Zellen sowie in den Zellen ohne die Desaminasen Cda und Dctd absolut quantifiziert. (Abb. 4.23)

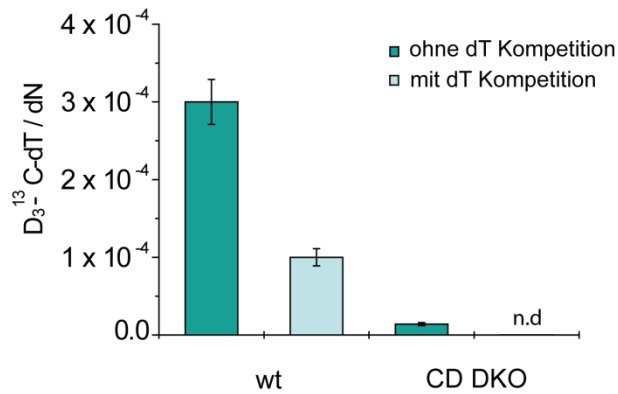


Abb. 4.23: Absolute Quantifizierung der Mengen an $D_3-^{13}C-dT$ in Wildtyp Zellen (wt) sowie in Zellen ohne die cytosolischen Desaminasen Cytosin - und Desoxycytidylat Desaminase (CD DKO) jeweils mit und ohne Konkurrenz durch natürliches Thymin.

Man erkennt, dass nach dem Ausschalten der Desaminasen Cda und Dctd zwar immer noch ein kleiner Rest an $D_3-^{13}C-dT$ detektierbar ist, der größte Anteil allerdings verschwindet, was für einen sehr großen Anteil an Desaminierung im löslichen Nukleosidpool spricht. Der verbleibende Rest an $D_3-^{13}C-dT$ kann durch Konkurrenz mit natürlichem dT weiter verringert werden. Das spricht für weiterhin vorhandene Desaminierung im löslichen Pool an Nukleosiden. Bei einer Konkurrenz mit natürlichem dT in CD DKO Zellen ist zwar immer noch ein Signal für $D_3-^{13}C-dT$ sichtbar, dieses liegt allerdings unter dem Quantifizierungslimit (*LOQ*). Somit könnte die verbleibende sehr geringe Menge an $D_3-^{13}C-dT$ zum Einen von weiteren unbekannten und daher nicht ausgeschalteten Desaminasen stammen, oder es könnte sich hierbei tatsächlich um den genomischen Anteil an Desaminierung von 5mdC zu dT handeln. An dieser Stelle muss also festgehalten werden, dass der Großteil des $D_3-^{13}C-dT$ durch Desaminierung im löslichen Nukleosidpool aus $D_3-^{13}C-5mdC$ / $D_3-^{13}C-5mdCMP$ durch die beiden Desaminasen Cda und Dctd generiert wird. Das $D_3-^{13}C-5mdC$ bzw. $D_3-^{13}C-5mdCMP$ gelangt vermutlich über Reparatur größerer DNA Fragmente in den löslichen Nukleosidpool. Daher wird dieser Vorgang im Folgenden als Umsatz von 5mdC beschrieben. Von einer Desaminierung wird im Gegensatz dazu nur noch gesprochen sofern es um die direkte enzymatische Umsetzung von 5mdC zu dT geht.

Wie bereits erwähnt könnte die Desaminierung von genomischem 5mdC durch Cytosin Desaminasen der AID/APOBEC Familie oder durch die *de novo* DNA Methyltransferasen Dnmt 3a/b erfolgen.^[9, 35, 36, 92, 101-103, 167] Um dies zu untersuchen wurden AID +/+ Zellen (AID wt) und AID -/- Zellen (AID KO) hinsichtlich ihrer Menge an gebildetem

genomischen D_3 - ^{13}C -dT sowie der Mengen an dU, 5mdC, 5hmdC und 5fdC untersucht. (Abb. 4.24)

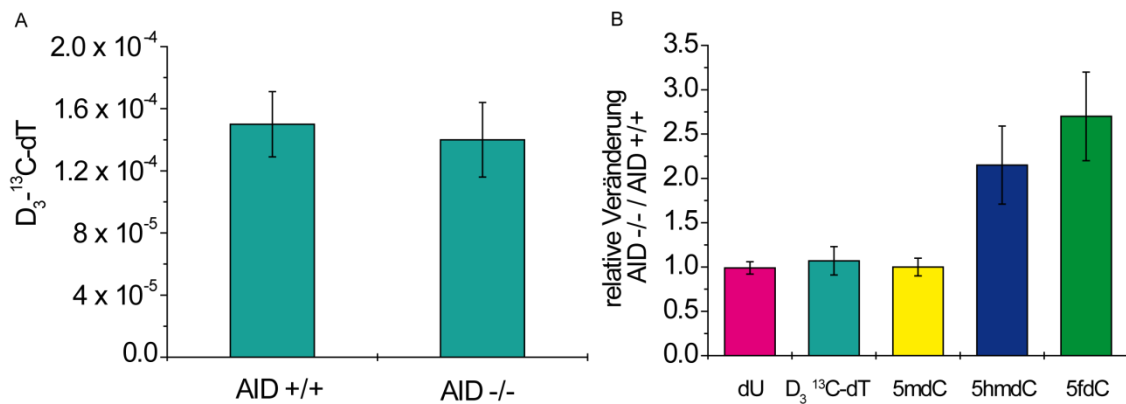


Abb. 4.24: A) Absolute Quantifizierung der Menge an D_3 - ^{13}C -dT in AID +/+ und AID -/- Zellen. B) Darstellung der relativen Veränderung der Level an dU, D_3 - ^{13}C -dT, 5mdC, 5hmdC und 5fdC in AID KO Zellen im Vergleich zu AID wt Zellen, N=3.

Man erkennt in Abbildung 4.24 A, dass in den AID -/- Zellen die Menge an D_3 - ^{13}C -dT etwa der Menge in den AID +/+ Zellen entspricht. Dies spricht gegen eine Beteiligung von AID an der Desaminierung. Hätte AID eine zentrale Rolle bei der Desaminierung von 5mdC zu dT, würde man im AID KO deutlich geringere Level an D_3 - ^{13}C -dT erwarten als im AID wt. Dennoch könnten die Mitglieder der AID/APOBEC Familie natürlich eine redundante Rolle bei der Desaminierung von genomischen 5mdC einnehmen. Bei einer Darstellung der relativen Veränderungen (Abb. 4.24 B) der Level an 5mdC, 5hmdC, 5fdC, dU und dT im AID KO im Vergleich zu den Wildtyp Zellen fällt auf, dass die Mengen an dU, 5mdC und D_3 - ^{13}C -dT nahezu unverändert bleiben, während sich die Mengen an 5hmdC und 5fdC im AID KO verdoppeln bis verdreifachen. Somit scheint AID zwar einen Einfluss auf die oxidierten Spezies 5hmdC und 5fdC zu haben, nicht aber auf 5mdC, dU und dT.

Auch die DNA Methyltransferasen Dnmt 3a und Dnmt 3b kommen als Desaminasen in Frage. Daher wurden die Mengen an D_3 - ^{13}C -dT im Dnmt wt im Vergleich zu einem Dnmt 3a und 3b *Knockout* untersucht. (Abb. 4.25)

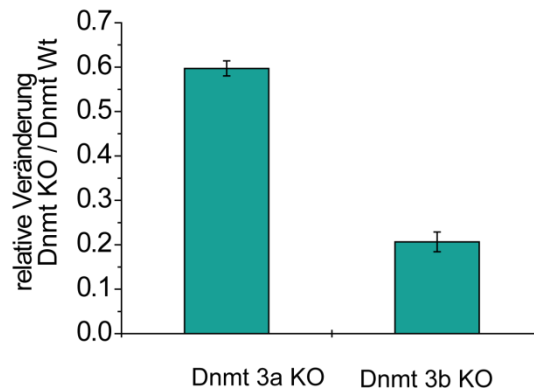


Abb. 4.25: Darstellung der relativen Veränderung der Level $D_3\text{-}^{13}\text{C-dT}$ im Dnmt 3a KO und Dnmt 3b KO im Vergleich zu Dnmt Wildtyp Zellen, N=3.

Interessanterweise sieht man hier bei Dnmt 3b einen deutlichen Effekt. Ein *Knockout* von Dnmt 3b scheint die Mengen an gebildetem $D_3\text{-}^{13}\text{C-dT}$ drastisch zu verringern. Ein *Knockout* von Dnmt 3a dagegen hat zwar einen Effekt auf die Menge an 5mdC, scheint jedoch kaum Auswirkung auf die Menge an $D_3\text{-}^{13}\text{C-dT}$ zu haben. Somit könnte Dnmt 3b direkt an der Desaminierung von 5mdC zu dT beteiligt sein. Ob die direkte Beteiligung sich in einer Funktion von Dnmt 3b als Desaminase von 5mdC äußert, oder ob Dnmt 3b indirekt beteiligt ist, indem es die Reparatur größerer DNA Fragmente mit auslöst oder steuert ist allerdings schwer zu beurteilen.

Als nächstes sollte die Frage geklärt werden in wie weit die Tet Enzyme an der Umsetzung von mC zu dT beteiligt sind oder anders gesagt, welchen Einfluss ein Fehlen der Tet Enzyme auf die Umsetzung von 5mdC zu dT hat. Zu diesem Zweck wurde die Menge an gebildetem $D_3\text{-}^{13}\text{C-dT}$ in Tet 1.3 Zellen, welche als Tet Wildtyp Referenz bezeichnet werden und in Tet Triple Knockout Zellen (Tet TKO) absolut quantifiziert. In Tet TKO Zellen sind die Enzyme Tet1, Tet2 und Tet3 ausgeschaltet, wodurch keine Oxidation von 5mdC zu 5hmdC oder 5fdC und 5cadC mehr möglich ist. Während der Differenzierung der Tet TKO Zellen konnte vermehrt die Entstehung toter Zellen beobachtet werden. Diese Ansammlung toter Zellen könnte zu Schwierigkeiten bei der absoluten Quantifizierung der $D_3\text{-}^{13}\text{C-dT}$ Werten führen, bzw. diese verfälschen. In verschiedenen Experimenten konnten sehr unterschiedliche Mengen an $D_3\text{-}^{13}\text{C-dT}$ im Tet TKO bestimmt werden, während der Level an $D_3\text{-}^{13}\text{C-dT}$ im Tet 1.3 Wildtyp annähernd konstant blieb. (Abb. 4.26)

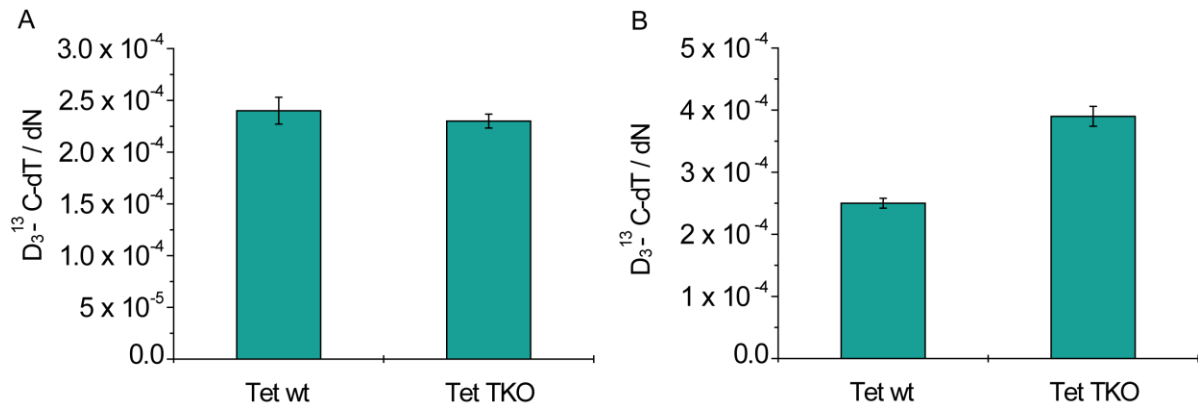


Abb. 4.26: Absolute Quantifizierung der Menge an $D_3\text{-}^{13}\text{C-dT}$ in Tet wt und Tet TKO Zellen. Unter A) sind gleichbleibende Mengen an $D_3\text{-}^{13}\text{C-dT}$ dargestellt, $N=3$. Unter B) Unterschiedliche Level an $D_3\text{-}^{13}\text{C-dT}$ im Tet wt und Tet TKO, $N=3$.

Bei allen Messungen blieben die Mengen an $D_3\text{-}^{13}\text{C-dT}$ im Tet TKO im Vergleich zum Tet Wildtyp entweder gleich oder wurden etwas mehr. In keinem der Experimente konnte für den Tet TKO ein niedrigerer Wert als für den Wildtyp detektiert werden. Dies spricht dafür, dass die Tet Enzyme für die Desaminierung keine große Rolle spielen. Die gelegentliche Steigerung der $D_3\text{-}^{13}\text{C-dT}$ Werte ließe sich möglicherweise mit einer gesteigerten Desaminierungsaktivität in Abwesenheit des alternativen Weges über Tet vermittelte Oxidation und Basenexzisionsreparatur erklären. Es könnte sich daher bei der Demethylierung über Desaminierung von 5mdC zu dT um einen *Backup* Weg handeln. Um diese These weiter zu stützen wurde ein *Puls-Chase*-Experiment durchgeführt. Bei diesem Experiment wurde zu einem bestimmten Zeitpunkt die Fütterung von $D_3\text{-}^{13}\text{C-SAM}$ eingestellt. Dadurch, dass kein zusätzliches $D_3\text{-}^{13}\text{C-5mdC}$ mehr gebildet werden kann, können die Veränderungen in den Mengen an $D_3\text{-}^{13}\text{C-5mdC}$ über einen Zeitraum von 12-50 h nach Beendigung der Fütterung beobachtet werden. Gleichzeitig wurde dieselbe Kultur mit schwermarkiertem dT ($^{13}\text{C}_{10}\text{-}^{15}\text{N}_2\text{-dT (dT+12)}$) gefüttert um die Zellproliferation zu verfolgen. (Abb. 4.27)

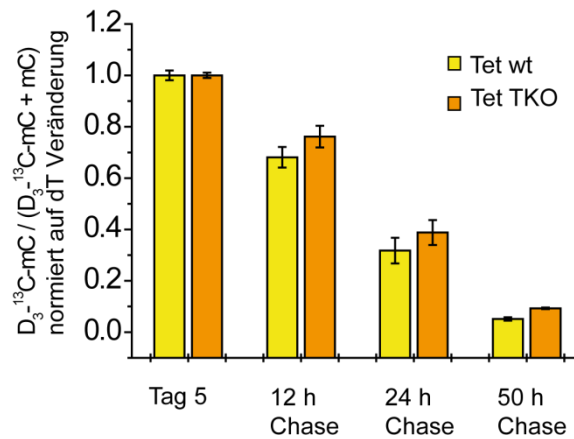


Abb. 4.27: Darstellung der relativen Veränderungen der Level an D_3 - ^{13}C -5mdC Level in Tet wt und Tet TKO Zellen an Tag 5 der Kultivierung und nachdem 12, 24 und 50 h nicht mehr mit D_3 - ^{13}C -SAM gefüttert wurde (=Chase). Die Werte wurden dabei auf die entsprechende Menge an dT+12 normiert um die Abnahme durch DNA Replikation vernachlässigen zu können.

In dieser Darstellungsweise erkennt man, dass es in den Tet Wildtyp Zellen, aber auch in der Tet TKO Zelllinie zu einer Abnahme an D_3 - ^{13}C -5mdC nach 12 h ohne Fütterung von D_3 - ^{13}C -SAM kommt. Zu einer weiteren Abnahme der relativen Mengen kommt es nach 24 bzw. 50 h ohne Fütterung. Diese Abnahme ist in den Tet wt Zellen stärker ausgeprägt, da hier durch die Tet Enzyme Tet 1, Tet 2 und Tet 3 zusätzlich D_3 - ^{13}C -5mdC zu D_3 - ^{13}C -5hmdC usw. oxidiert und damit entfernt wird. Aber auch in den Tet TKO Zellen ist eine deutliche Abnahme der relativen Menge an D_3 - ^{13}C -5mdC detektierbar. An dieser Stelle muss erwähnt werden, dass in diesem Experiment zusätzlich isotopenmarkiertes dT+12 zugegeben wurde. Auch die Fütterung von dT+12 wurde an Tag 5 eingestellt und die jeweiligen Mengen 12, 24 und 50 h nach der Fütterung bestimmt. Da man davon ausgehen kann, das dT+12 nur durch Replikation über die Zeit verdünnt wird, kann man die Menge, die durch Verdünnung verloren geht in die Berechnung der relativen Veränderungen der D_3 - ^{13}C -5mdC Werte mit einbeziehen. Die D_3 - ^{13}C -5mdC Level wurden daher auf die dT+12 Level normiert und so die replikationsabhängige Verdünnung heraus gerechnet. Bei der in Abb. 4.27 dargestellten relativen Veränderung von D_3 - ^{13}C -5mdC muss die Replikation daher nicht mehr berücksichtigt werden. Somit ist klar, dass es sich bei der Abnahme in den Tet TKO Zellen um eine replikations- und oxidationsunabhängige Reduktion der D_3 - ^{13}C -5mdC Mengen handelt. Dabei könnte es sich um den Anteil der D_3 - ^{13}C -5mdC Abnahme handeln, welche durch Umsatz von D_3 - ^{13}C -5mdC zu D_3 - ^{13}C -dT zu Stande kommt.

Im Folgenden sollte der Frage nachgegangen werden, welche Reparaturwege am Umsatz von 5mdC zu dT beteiligt sind. Normalerweise erkennen DNA-Reparatursysteme bestimmte DNA Schäden und reparieren diese gezielt. Durch eine direkte Desaminierung von 5mdC zu dT würden T:G Fehlpaarungen entstehen, die normalerweise von Tdg oder Mbd4 erkannt und mit Hilfe der Basenexzisionsreparatur behoben werden. Aus der Untersuchung der *Knockouts* der cytosolischen Desaminasen und der T Kompetition haben sich Hinweise auf einen geringen Anteil an genomischen Desaminierung ergeben. Daher sollte im Folgenden der Einfluss der Glykosylasen Tdg und Mbd4 auf die Entfernung von D_3 - ^{13}C -dT untersucht werden. Hierzu wurden die Level an D_3 - ^{13}C -dT in Tdg $-/-$ mESC im Vergleich zu Tdg $+/-$ Zellen bestimmt. (Abb. 4.28)

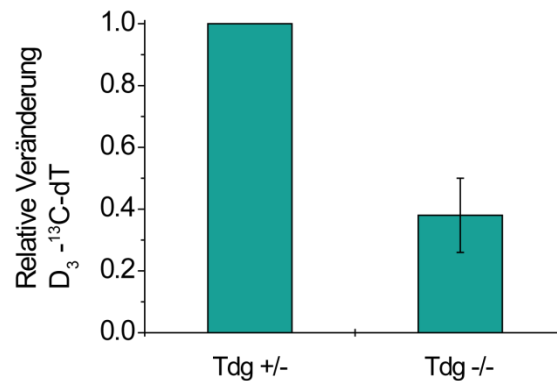


Abb. 4.28: Darstellung der relativen Veränderung der Level D_3 - ^{13}C -dT in Tdg $-/-$ Zellen im Vergleich zu Tdg $+/-$ Zellen, N=8.

In den Tdg $-/-$ mESC Zellen wurde etwa die Hälfte an D_3 - ^{13}C -dT im Vergleich zu Tdg $+/-$ Zellen detektiert. Diese Daten konnten achtmal reproduziert werden, was dieses Ergebnis sehr verlässlich macht. Dies ist ein überraschendes Ergebnis, denn eigentlich sollte davon ausgegangen werden, dass die detektierte Menge an D_3 - ^{13}C -dT größer werden sollte im Vergleich zu Tdg $+/-$ Zellen wenn man die Glykosylase, welche den Schaden ausschneidet, eliminiert. Allerdings konnte bereits durch den *Knockout* der cytosolischen Desaminasen sowie durch ein Kompetitionsexperiment mit natürlichem dT gezeigt werden, dass der Anteil an genomischer Desaminierung nur sehr klein ist. Der meiste Umsatz an 5mdC zu dT findet im löslichen Nukleosidpool statt. Daher wäre auch kein großer Effekt der Basenexzisionsreparatur zu erwarten. Dennoch sollte keine

Abnahme zu verzeichnen sein.

Möglicherweise übernimmt Mbd4 oder die Glykosylasen NEIL1 und NEIL2 bei Abwesenheit von Tdg dessen Aufgabe, sodass ein Fehlen von Tdg nicht so stark ins Gewicht fallen würde. Im Folgenden wäre es daher sinnvoll eine Mbd4 *Knockout* Zelllinie, eine NEIL1 und 2 *Knockout* Zelllinie sowie einen Tdg und Mbd4 *Doppelknockout* hinsichtlich ihrer Mengen an D₃-¹³C-dT zu untersuchen.

Nach Reparatur durch die Glykosylasen Tdg oder Mbd4 könnte ein Reparaturweg aktiviert werden, welcher 5mdC (markiert und unmarkiert) zusammen mit größeren DNA-Fragmenten ausschneidet. Dadurch würde vermehrt 5mdC in den löslichen Pool an Nukleosiden gelangen und dort von den cytosolischen Desaminasen Cda und Dctd desaminiert werden. Eine Re-inkorporierung der gebildeten dT-Nukleoside würde dann zu den ungewöhnlich hohen Mengen an markiertem dT im Genom führen. Die Erkennung des Schadens durch Tdg oder eine andere Glykosylasen wäre dann gewissermaßen nur das Signal welches zur weiteren Prozessierung des Schadens durch andere alternative Reparaturwege führt.

Im Folgenden sollte der Frage nachgegangen werden ob es sich bei diesem alternativen Reparaturweg um Nukleotidexzisionsreparatur (NER) handeln könnte. Im Unterschied zur Basenexzisionsreparatur werden von der Nukleotidexzisionsreparatur größere DNA Fragmente von 24 bis 32 Nukleosiden entfernt.^[174] Daher könnte durch NER größere Mengen an D₃-¹³C-5mdC in den löslichen Nukleosidpool gelangen und nach Desaminierung und Re-inkorporierung in Form von markiertem D₃-¹³C-dT in der DNA gefunden werden. Bei der Nukleotidexzisionsreparatur erfolgt die Erkennung eines Schadens in Eukaryoten entweder durch globale Genomreparatur (GGR) wenn sich der Schaden in transkriptionsinaktiven Bereichen der DNA befindet oder durch transkriptionsgekoppelte Reparatur (TCR) bei der die RNA-Polymerase durch Anwesenheit eines Schadens blockiert wird und so den Reparaturprozess einleitet.^[175-178] Am TCR-NER ist das Enzym Csb beteiligt. Csb unterstützt zusammen mit Csa das Entfernen der RNA Polymerase II, welche die geschädigte DNA blockiert und ermöglicht somit die Reparatur. Nach der Erkennung des Schadens erfolgt die Entwindung der DNA und die Prozessierung des Schadens. Diese Prozesse verlaufen im GGR-NER und im TCR-NER identisch ab. Das Protein XPA ist maßgeblich an diesen Schritten beteiligt. Nach der Entfernung des geschädigten Bereiches der DNA, wird die DNA an dieser Stelle durch Polymerasen neu synthetisiert und anschließend ligiert.

Um zu überprüfen ob die Nukleotidexzisionsreparatur an der Reparatur von $D_3\text{-}^{13}\text{C-dT}$ beteiligt ist, wurden *Knockout*-Zelllinien der zentralen Enzyme Csb und Xpa untersucht. (Abb. 4.29)

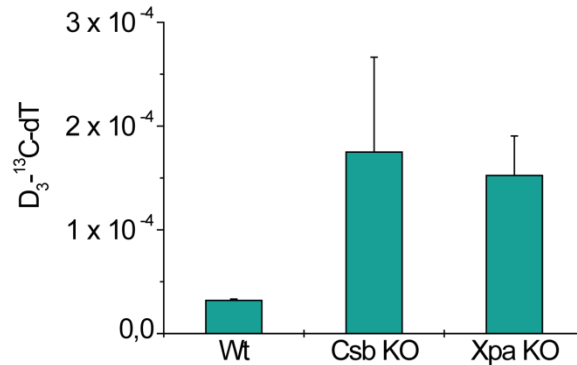


Abb. 4.29: Absolute Quantifizierung der $D_3\text{-}^{13}\text{C-dT}$ - Mengen in Wildtyp Zellen sowie Zelllinien in denen Csb und Xpa, zentrale Enzyme des NER, ausgeschaltet sind.

Man erkennt, dass die Level an quantifiziertem $D_3\text{-}^{13}\text{C-dT}$ in beiden *Knockouts* ansteigen. Eigentlich wäre zu erwarten gewesen, dass die Menge an gebildetem $D_3\text{-}^{13}\text{C-dT}$ abnimmt wenn man den Reparaturwegs ausschaltet. Somit scheint NER entweder nicht beteiligt zu sein oder ein anderes Reparatursystem übernimmt diese Aufgabe und setzt 5mdC wesentlich effizienter zu dT um. Bei diesem anderen Reparatursystem könnte es sich um den Weg der *Mismatch*-Reparatur handeln.

Während der Replikation können durch Oxidation, Desaminierung und Methylierung Basenfehlpaarungen entstehen. Diese fehlgepaarten Basen werden durch die *Mismatch*-Reparatur (MMR) behoben.^[179] In Eukaryoten sind die Enzyme Mlh1, Mlh3, Msh2, Msh3, Msh6, Pms1 und Pms2 an der *Mismatch*-Reparatur beteiligt. Das Enzym Mlh1 ist dabei zusammen mit Pms2 am Ausschneiden des Schadens beteiligt, Msh2 scheint dagegen bei der Erkennung des Schadens eine Rolle zu spielen.^[180, 181] Die Auswirkung der *Mismatch*-Reparatur auf die Desaminierung von 5mdC zu dT wurde anhand von Mlh1 und Msh2 *Knockouts* untersucht. (Abb. 4.30)

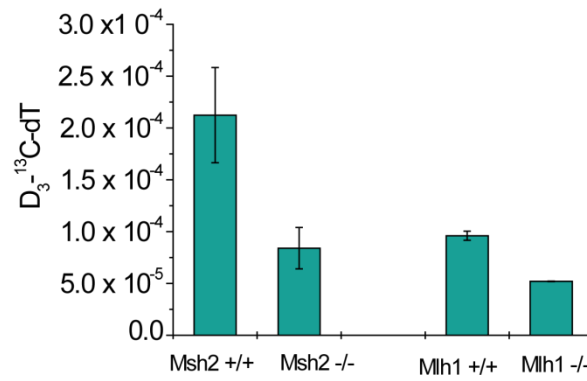


Abb. 4.30: Absolute Quantifizierung der $D_3\text{-}^{13}\text{C-dT}$ - Mengen in Msh2 und Mlh1 Wildtyp IB10 Zellen sowie in Zelllinien denen Msh2 und Mlh1, zentrale Enzyme des *Mismatch*-Reparatur, fehlen.

Man erkennt, dass sowohl in Msh2 *Knockout* Zellen als auch in Mlh1 *Knockout* Zellen die Mengen an $D_3\text{-}^{13}\text{C-dT}$ im Vergleich zum jeweiligen Wildtyp etwa zur Hälfte reduziert ist. Somit scheint das Fehlen der für die Erkennung bzw. der Behebung des Schadens zuständigen Enzyme der *Mismatch*-Reparatur eine große Auswirkung auf die Mengen an $D_3\text{-}^{13}\text{C-dT}$ im Genom zu haben. Diese Daten müssen jedoch noch ein weiteres Mal reproduziert werden und können daher nur als erster Hinweis gesehen werden.

Ein weiterer Hinweis ergibt sich aus einem *Puls-Chase*-Experiment mit Mlh1 wt und Mlh1 KO Zellen. Auch hier wurden die Zellen bis Tag 5 der Kultivierung mit $D_3\text{-}^{13}\text{C-SAM}$ gefüttert und die Fütterung anschließend eingestellt. Die Zellen wurden 12 h, 24 h und 50 h nach Beendigung der Fütterung geerntet und die relative Veränderung an $D_3\text{-}^{13}\text{C-5mdC}$ normiert auf die Menge an dT+12 bestimmt. (Abb. 4.31)

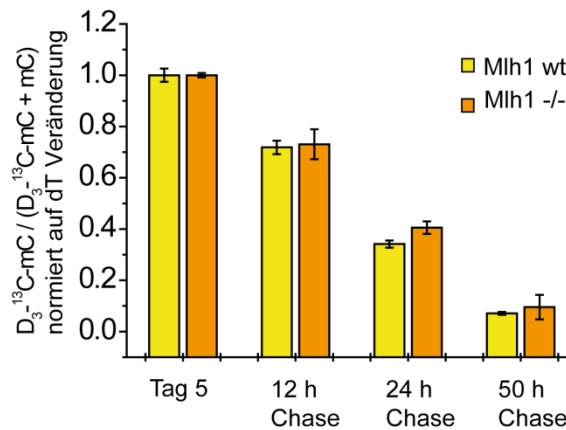


Abb. 4.31: Darstellung der relativen Veränderungen der Level an $D_3\text{-}^{13}\text{C}$ -5mdC Level in Mlh1 wt und Mlh1 KO Zellen an Tag 5 der Kultivierung und nachdem 12, 24 und 50 h nicht mehr mit $D_3\text{-}^{13}\text{C}$ -SAM gefüttert wurde. Die Werte wurden dabei auf $^{13}\text{C}_{10}\text{-}^{15}\text{N}_2\text{-dT}$ normiert um die Abnahme durch DNA Replikation vernachlässigen zu können.

Als Referenz zur Bestimmung der replikationsabhängigen Verdünnung von $D_3\text{-}^{13}\text{C}$ -5mdC diente hier wie schon zuvor $^{13}\text{C}_{10}\text{-}^{15}\text{N}_2\text{-dT}$ als Referenz. Somit sind die gezeigten relativen Abnahmen an $D_3\text{-}^{13}\text{C}$ -5mdC unabhängig von der Verdünnung durch Replikation. Man erkennt sowohl bei Mlh1 wt als auch bei Mlh1 -/- Zellen eine Abnahme der relativen Menge an $D_3\text{-}^{13}\text{C}$ -5mdC nachdem die Fütterung von $D_3\text{-}^{13}\text{C}$ -SAM eingestellt wurde. Die Abnahme ist bei den Wildtyp Zellen etwas stärker ausgeprägt als bei der Mlh1 -/- Zelllinie. Somit scheint Mlh1 und damit die *Mismatch*-Reparatur einen Einfluss auf den Umsatz von $D_3\text{-}^{13}\text{C}$ -5mdC zu haben. Demnach könnte die *Mismatch*-Reparatur tatsächlich derjenige Reparaturweg sein, welcher größere DNA Fragmente ausschneidet und wodurch 5mdC in den Pool löslicher Nukleoside gelangt. Durch Desaminierung mit Hilfe der Desaminasen des löslichen Nukleosidpools entsteht schließlich dT welches anschließend wieder in DNA inkorporiert wird. Dies würde die unnatürlich großen Mengen an markiertem $D_3\text{-}^{13}\text{C}$ -dT im Genom erklären.

Zusammenfassend kann man sagen, dass es Hinweise auf genomische Desaminierung von 5mdC zu dT gibt. Durch umfangreiche Experimente konnte gezeigt werden, dass der größte Anteil an durch Desaminierung gebildetem dT durch das Ausschneiden größerer DNA Fragmente und anschließender Desaminierung durch Desaminasen im löslichen Nukleosidpool erfolgt. Im weiteren Verlauf kommt es zu Reinkorporierung des dT in die DNA. Das Ausschneiden der größeren DNA Fragmente scheint nicht durch Nukleotidexzisionsreparatur zu erfolgen, es gibt jedoch Hinweise, dass die *Mismatch-Reparatur* an der Prozessierung der Schäden beteiligt ist.

Dabei stellt sich natürlich auch die Frage in wie weit das recycling über tote Zellen eine Rolle spielt und die Mengen an $D_3\text{-}^{13}\text{C}\text{-dT}$ beeinflusst. Zu diesem Zweck wurden Messungen der Nukleoside im Medium normal wachsender Kulturen wie der Tet 1.3 Zelllinie mit den Nukleosiden im Medium einer Kultur mit vermehrtem Zelltod, wie der Tet TKO Zelllinie, verglichen. In ersten Experimenten konnte kein Unterschied in der Menge an $D_3\text{-}^{13}\text{C}\text{-dT}$ detektiert werden, um einen Einfluss von toten Zellen auf die Mengen an $D_3\text{-}^{13}\text{C}\text{-dT}$ völlig ausschließen zu können sollte dieses Experiment allerdings mit größeren Mengen an Medium wiederholt und die Detektion von $D_3\text{-}^{13}\text{C}\text{-dT}$ optimiert werden.

In wie weit Basenexzisionsreparatur eine Rolle spielt ist immer noch fraglich. Hierzu sollte ein Tdg und Mbd4 Doppelknockout untersucht werden.

Während die Cytosin Desaminase AID keinen Einfluss auf globale Level an $D_3\text{-}^{13}\text{C}\text{-dT}$ zu haben scheint, zeigt die DNA-Methyltransferase Dnmt 3b einen deutlichen Effekt. Dessen Abwesenheit verringert die Menge an $D_3\text{-}^{13}\text{C}\text{-dT}$ wesentlich. Somit scheint Dnmt 3b an der Desaminierung beteiligt zu sein. Ob diese Beteiligung sich in einer Funktion von Dnmt 3b als Desaminase von 5mdC äußert, oder ob Dnmt 3b indirekt beteiligt ist, indem es die Reparatur größerer DNA Fragmente mit auslöst oder steuert, ist allerdings schwer zu beurteilen.

Die Tet Enzyme dagegen scheinen keinen Einfluss auf die Desaminierung von 5mdC zu dT zu haben. Es könnte sich allerdings bei der Desaminierung um einen *Backup* Weg für aktive DNA Demethylierung handeln, wenn der oxidative Weg durch das Fehlen der Tet Enzyme nicht mehr möglich ist. Ein Hinweis hierauf ergeben Experimente zum Umsatz an 5mdC in An- und Abwesenheit der Tet Enzyme im Vergleich zum Umsatz durch Replikation. Hier ergaben sich deutliche Hinweise auf einen zusätzlichen Umsatz von 5mdC der weder replikations- noch oxidationsabhängig ist.

4.3.4 Zusammenfassung und Projektbeiträge

Die Entwicklung der Methode zur sensitiven Quantifizierung von D_3 - ^{13}C -dT wurde in enger Zusammenarbeit mit Olesea Kosmatchev durchgeführt. Zudem wurden die Verdaubedingungen und die Verwendung von Desaminaseinhibitoren so optimiert, dass zuverlässig quantifiziert werden konnte. Die Zellkulturversuche und die Isolation der genomischen DNA wurden von Fabio Spada und Angie Kirchner durchgeführt. Die Generierung der KO-Zelllinien Cda und Dctd Doppelknockout, Tdg und Mbd4 *Doppelknockout* sowie dem Mlh1 KO erfolgte durch Angie Kirchner.

Der enzymatische Verdau der Proben, die Messung, Auswertung und Interpretation der Daten wurden von Fabio Spada, Olesea Kosmatchev und mir vorgenommen.

4.3.5 Material und Methoden

DNA Isolation

Die Isolation der DNA erfolgte nach dem unter 3.3 beschriebenen Protokoll.

Verdau

Der Verdau der genomischen DNA (4-7 µg) erfolgte in 50 µL ddH₂O. Die DNA wurde in 35 µL ddH₂O vorgelegt. Für den ersten Verdauschnitt wurden 7.5 µL wässrige Lösung zugegeben. Diese enthielt 480 µM ZnSO₄, 42 U Nuklease S1 (Sigma-Aldrich), 5 U Antarctic Phosphatase, 100 µM THU und eine spezifische Menge an gelabelten internen Standards. Die Mischung wurde bei 37°C für 3 h inkubiert. Für den zweiten Verdauschnitt wurden ebenfalls 7.5 µL wässrige Lösung zugegeben. Diese enthielt 520 µM [Na]₂-EDTA-Lösung und 0.2 U Snake Venom Phosphodiesterase (*Crotalus adamanteus*, USB Corporation) und wurde bei 37°C für 3h inkubiert. Die Proben wurden bis zum Tag der Analyse bei -20°C aufbewahrt. Vor der UHPLC-MS/MS Analyse wurden die Proben mit AcroPrep Advance 96 Filterplatten 0.2 µM Supor (Pall Life Sciences) filtriert.

LC-MS/MS-Methode

Die quantitative LC-ESI-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte wie unter 4.1.5 beschrieben mit einem 1290 UHPLC-System von *Agilent*, welches über einen UV Detektor verfügt und mit einem 6490 Triplequadrupol Massenspektrometer von *Agilent* verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8-Säule (*Agilent*, 2.7 µm, 2.1 mm × 150 mm) bei 30°C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0.0085% (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0.35 ml/min: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 7.2 min; 5 → 80% MeCN; 7.2 → 10.5 min; 80% MeCN; 10.5 → 11.3 min; 80 → 0% MeCN; 11.3 → 13 min; 0% MeCN.

Die Quellparameter waren dieselben wie unter 4.1.5. Die für die jeweiligen Analyten spezifischen Parameter sind in Tabelle 4.3 zusammengefasst.

Tabelle 4.3: Substanz abhängige LC/MS-MS Parameter zur Quantifizierung der Produkte der D₂-¹³C-S-Adenosylmethionin Fütterung sowie der natürlichen Nukleoside in genomischer DNA

Verbindung	Precursor Ion (<i>m/z</i>)	MS1 Auflösung	Produkt Ion (<i>m/z</i>)	MS2 Auflösung	Dwell time [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1.5–4.0 min								
[¹⁵ N ₂]-cadC	274.08	Wide	158.04	wide	40	5	5	Positive
[¹³ C]-cadC	273.09	Wide	157.04	wide	40	5	5	Positive
cadC	272.09	Wide	156.04	wide	40	5	5	Positive
[D ₂ , ¹⁵ N ₂]-hmdC	262.11	Wide	146.06	Wide	40	27	1	Positive
[D ₂ , ¹³ C]-hmdC	261.11	Wide	145.06	Wide	40	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	40	27	1	Positive
[¹³ C ₁₀ , ¹⁵ N ₂]-mdC	247.13	Wide	131.09	Wide	30	60	1	Positive
[D ₃ , ¹³ C]-mdC	246.13	Wide	130.09	Wide	30	60	1	Positive
mdC	242.11	Wide	126.07	Wide	30	60	1	Positive
dC	228.1	Wide	112.1	Wide	40	1	3	Positive
Zeitsegment 4.0–6.0 min								
[D ₂ , ¹³ C]-hmdU	260.08	Wide	217.07	Wide	50	7	5	Negative
hmdU	257.08	Wide	214.07	Wide	50	7	5	Negative
[D ₁ , ¹³ C]-fdU	257.06	Wide	214.05	Wide	50	6	5	Negative
fdU	255.06	Wide	212.06	Wide	50	6	5	Negative
[¹⁵ N ₂]-dU	229.06	Wide	185.06	Wide	50	5	5	Negative
dU	227.07	Wide	184.06	Wide	50	5	5	Negative
Zeitsegment 6.0–9.0 min								
[¹⁵ N ₅]-8-oxo-dG	289.08	Wide	173.04	Wide	80	9	7	Positive
8-oxo-dG	284.1	Wide	168.05	Wide	80	9	7	Positive
[¹⁵ N ₂]-fdC	258.09	Wide	142.04	Wide	80	5	5	Positive
fdC	256.09	Wide	140.05	Wide	80	5	5	Positive
[D ₃ , ¹³ C]-dT	247.1	Wide	131.05	Wide	30	40	3	Positive
dT+12	255.1	Wide	134.05	Wide	30	40	3	Positive
dT	243.1	Wide	127.05	Wide	30	40	3	Positive

Validierung der LC-MS Methode

Die Validierung erfolgte wie unter 4.1.5 beschrieben.

Eichgeraden:

Die für die Quantifizierung notwendigen Eichgeraden der gelabelten und korrespondierenden ungelabelten synthetisierten Nukleoside wurden bei 5 unterschiedlichen Konzentrationen aufgenommen. Die lineare Regression wurde unter Verwendung von Origin 6.0 (Microcal) durchgeführt. Hierfür wurde das Verhältnis der Flächen (A/A^*) unter dem Signal für das unmarkierte Nukleosid durch den internen Standard (*) gegen das Verhältnis der Stoffmengen (n/n^*) des unmarkierten Nukleosides durch den internen Standard aufgetragen. Für jede Konzentration wurde ein Mittelwert aus drei unabhängigen Messungen gebildet. (Abb. 4.32)

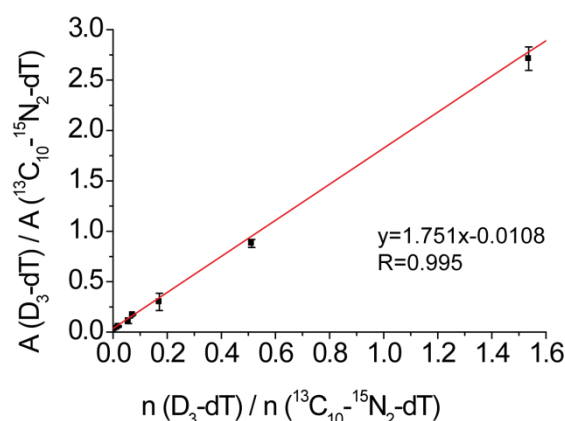


Abb. 4.32: Masseneichgeraden für die Nukleoside D_3-dT und $^{13}C_{10}-^{15}N_2-dT$

Vorausgesetzt wurden eine Genauigkeit von 80-120% sowie eine Präzision von weniger als 20 RSD. Die Genauigkeit wurde überprüft durch Berechnung der Stoffmengen durch das mit Hilfe der Eichgerade erhaltene A/A^* Verhältnis der Analyten. Die Genauigkeit wurde definiert als das Verhältnis aus eingesetzter Menge an Substanz durch die berechnete Menge an Substanz in Prozent. Diese Genauigkeit sollte zwischen 80 und 120% liegen.

Für die Präzision mussten bei technischen Replikaten die Standardabweichung (RSD) unter 20% liegen. Das untere Quantifizierungslimit (*lower limit of quantification*, LLOQ) wurde als niedrigste Konzentration definiert die grade noch die Anforderungen der Genauigkeit und Präzision erfüllt.

Die anderen Eichgeraden finden sich unter 3.3.

Auswertung

Die Auswertung erfolgte wie unter 3.3 beschrieben mit dem Programm Mass Hunter und Excel, die Darstellung der Daten erfolgte mit Origin und Adobe Illustrator.

4.4 Interaktoren und Biologische Funktion von ms²i⁶A

4.4.1 Prolog

Im Gegensatz zu DNA, welche als relativ statischer und wenig modifizierter Informationsspeicher gilt, finden sich in RNA sehr zahlreiche und zum Teil komplexe Modifikationen der Nukleoside. Die verschiedenen RNA Spezies dienen als Adaptermoleküle für die Proteinbiosynthese, regulieren die Genexpression, transportieren und übersetzen die genetische Information und dienen als Katalysatoren. So ist es nicht verwunderlich, dass in allen RNA Spezies bis heute über 100 unterschiedliche Modifikationen gefunden wurden.^[124, 125] Diese finden sich an allen vier kanonischen RNA-Basen und reichen z.B. im Falle des Adenosins von einfachen Methylierungen wie in *N*⁶-Methyladenosin (m⁶A) über komplexere Modifikationen wie *N*⁶-Threonylcarbamoyladenosin (t⁶A) oder *N*⁶-Isopentenyladenosin (i⁶A) bis hin zu sogenannten Hypermodifikationen wie 2-Methylthio-*N*⁶-isopentenyladenosin (ms²i⁶A).^[124] Das hypermodifizierte ms²i⁶A wird in zwei Stufen aus Adenosin gebildet. (Abb. 4.33)

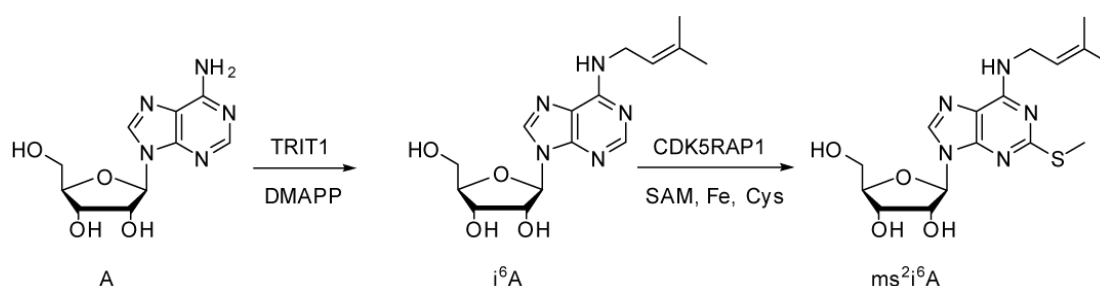


Abb. 4.33: Humaner Biosyntheseweg von i⁶A durch TRIT1 und DMAPP und ms²i⁶A durch das Enzym CDK5RAP1 mit Hilfe von SAM, Eisen III und Cystein.

In Menschen erfolgt dies durch eine Isopentenylierung von Adenosin durch das Enzym TRIT1,^[141] einem Homolog zum bakteriellen Mia A und einer anschließenden Thiomethylierung durch CDK5RAP1,^[147, 148] einem Homolog zum bakteriellen MiaB. An der Position 37 im Antikodonloop der tRNA übernimmt ms²i⁶A die Aufgabe der Stabilisierung von Kodon-Antikodon Wechselwirkungen.^[153]

Die Entdeckung geringer Mengen an ms²i⁶A in Poly-A-RNA^[147] sowie das Vorhandensein einer cytosolischen Variante des CDK5RAP1, die außerhalb der Mitochondrien arbeitet,^[149] lassen auf eine Funktion hypermodifizierter RNA Modifikation außerhalb von tRNA und der Mitochondrien schließen.

Im Folgenden soll die biologische Funktion von ms²i⁶A außerhalb der tRNA näher beleuchtet werden.

4.4.2 Ergebnisse und Diskussion

Um Aufschlüsse über die biologische Funktion von ms^2i^6A zu erhalten, sollten potentielle Interaktionspartner mit Hilfe eines SILAC (*stable isotope labeling by amino acids in cell culture*) Experimentes identifiziert werden. Der SILAC-Pulldown wurde analog zu früheren SILAC Experimenten in einer Kooperation mit der Gruppe von Michiel Vermeulen durchgeführt.^[91] Die benötigten RNA Sequenzen wurden jeweils mit und ohne die Modifikation ms^2i^6A synthetisiert. Als Sequenz wurde der $tRNA^{Phe}$ Anticodonloop verwendet, da die Modifikation ms^2i^6A in einem bekannten Sequenzkontext eingebaut werden sollte. Für das SILAC Experiment wurden der unmodifizierte sowie auch der modifizierte RNA-Strang jeweils in schweratommarkierten sowie in unmarkierten Zelllysaten inkubiert. Die Isotopenmarkierung erfolgt durch Verwendung der mit stabilen Isotopen markierten Aminosäuren Lysin⁸ und Arginin¹⁰ im Medium der Zellen. Die isotopenmarkierten Aminosäuren werden in die Proteine der Zellen inkorporiert und dienen anschließend der Unterscheidung unterschiedlicher Versuchsbedingungen. Dadurch, dass sich die bioisoteren Aminosäuren in der Massenspektrometrie klar unterscheiden lassen, ist das Mischen unterschiedlich behandelter Proben vor der Analyse möglich. (Abb. 4.34)

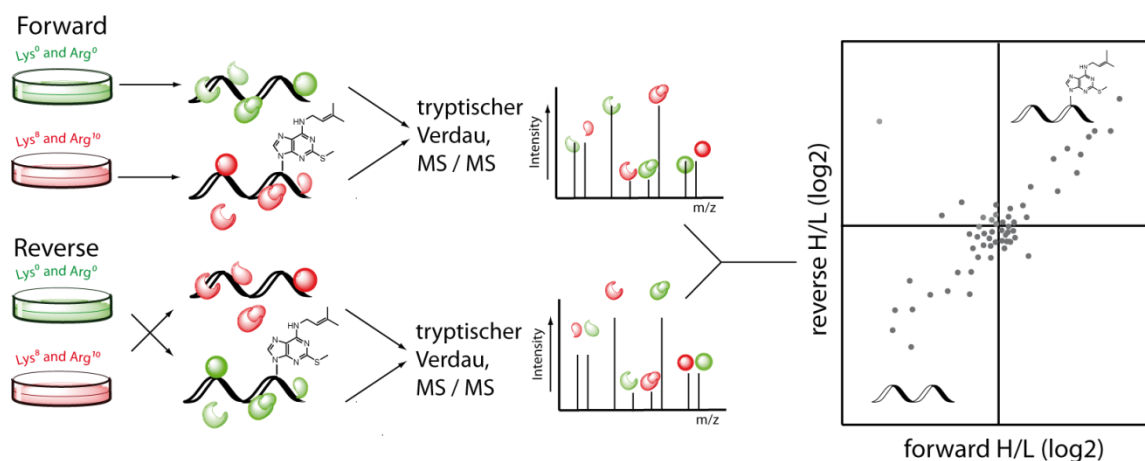


Abb. 4.34: Schematischer Ablauf eines SILAC Experimentes. Die Markierung der Proteine erfolgt durch die schwerisotopenmarkierte Aminosäuren Lysin⁸ und Arginin¹⁰. Durch die Markierung ist eine simultane Analyse sowie ein entgegengesetztes Kontrollexperiment möglich.

Die Lysate von HeLa Zellen wurden für die SILAC Experimente in cytosolisches und nukleares Lysat aufgetrennt. Während im Pulldown mit nuklearem HeLa-Lysat kaum Proteine angereichert werden konnten, wurden im *Pulldown* mit cytosolischem Lysat zahlreiche Proteine mit hoher Signifikanz von ms^2i^6A angezogen (blau markiert) oder abgestoßen (rot markiert). (Abb. 4.35)

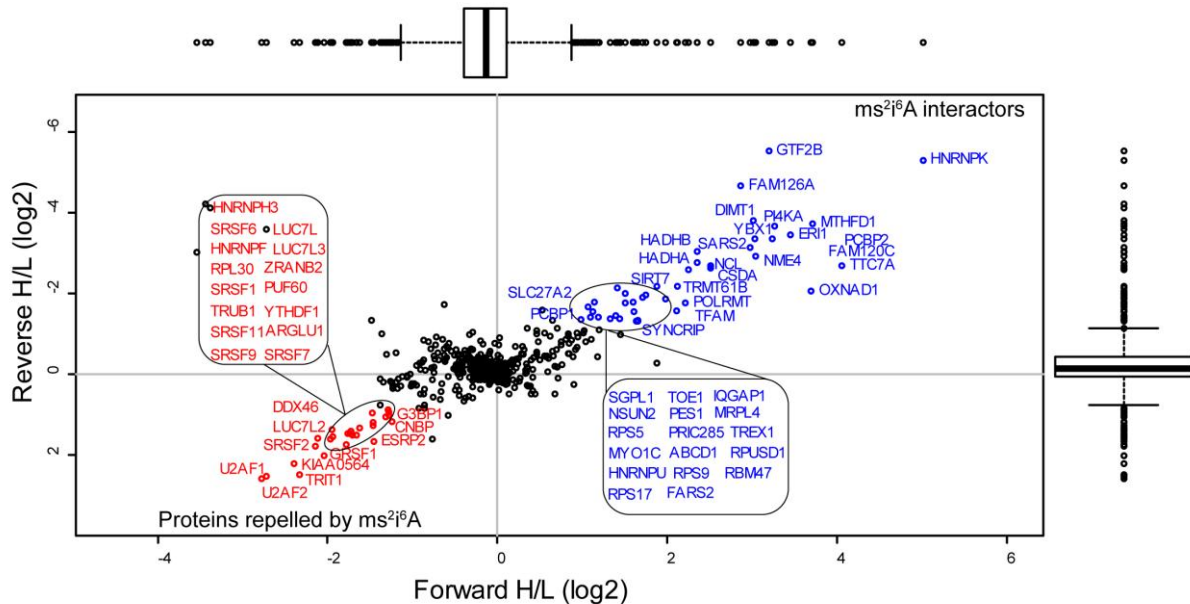


Abb. 4.35: Scatterplot des SILAC-Pulldowns mit ms^2i^6A enthaltender $tRNA^{Phe}$ und unmodifizierter $tRNA^{Phe}$. In blau dargestellt sind Proteine, die durch die Modifikation ms^2i^6A angezogen werden, rot markiert sind von ms^2i^6A abgestoßene Proteine.

Unter den am stärksten von ms^2i^6A angereicherten Proteinen findet sich das *heterogeneous nuclear ribonucleoprotein K* (HNRNPK). Dieses gehört zu den HNRNPs, einer Familie von RNA bindenden Proteinen die Komplexe mit *heterogeneous nuclear* RNA (hnRNA) bilden. Die HNRNPs sind mit pre-mRNAs im Nukleus verbunden und scheinen die Prozessierung von pre-mRNA, sowie auch den mRNA Metabolismus und Transport zu beeinflussen.^[182, 183] Man findet alle HNRNPs im Nukleus, einige wie auch HNRNPK können aber sowohl im Nukleus als auch im Cytoplasma lokalisiert sein.^[184]

Die HNRNPs verfügen über mehrere Wiederholungen von RNA-bindenden Domänen, den *KH-Repeats* und es konnte gezeigt werden, dass HNRNPK sowohl in die Regulation von alternativem Spleißen, als auch in die Stummschaltung von speziellen mRNAs involviert ist.^[185, 186]

Es scheint zudem ein Antagonist des im SILAC pulldown durch ms^2i^6A abgestoßenen Spleißing Faktoren U2AF1/2 zu sein. Es konnte gezeigt werden, dass HNRNPK sehr

spezifisch an eine RNA-Sequenz des *synaptosomal-associated protein 25* (Snap25) Genes bindet.^[187] Dieses Gen ist zuständig für das durch Forskolin inhibierte Spleißen während der neuronalen Differenzierung von pheochromocytoma PC12 Rattenzellen. Die Bindung von HNRNPK an diese Sequenz verdrängt U2AF2, welches ein essentieller Teil des frühen Spleiceosomes ist. Dadurch werden spezielle Exons angereichert, die mit neuronalen Krankheiten in Verbindung gebracht werden.^[187] Somit könnte auch in diesem Zusammenhang U2AF2 als Antagonist von HNRNPK auftreten.

Unter den ms²i⁶A Bindern finden sich noch mehr Proteine von denen bereits bekannt ist, dass sie mit HNRNPK im Komplex auftreten. So ist auch *Y-box-binding protein 1* (YBX1), welches unter den Bindern auftaucht, bereits früher mit Hilfe von *two-hybrid* Screening als Bindungspartner von HNRNPK identifiziert worden.^[188] Auch von Nukleolin^[189] und von einigen Mitgliedern der HNRNP-Familie wie PCBP1, PCBP2 und SYNCRIP ist eine Interaktion mit HNRNPK bekannt.^[185, 190] Es verwundert also nicht, dass sich all diese Proteine unter den von ms²i⁶A gebundenen bzw. abgestoßenen Proteinen befinden. Es ist wahrscheinlich, dass auch hier einige Proteine im Komplex gebunden haben. Der generelle Transkriptionsfaktor 2B (GTF2B), ein weiteres durch ms²i⁶A angereichertes Protein wurde dagegen in keinem der Komplexe um HNRNPK gefunden. Da es sich um einen Transkriptionsfaktor handelt, könnte dies auf eine weitere Funktion der Modifikation ms²i⁶A neben der Regulation des Spleißens hindeuten. Auch auf die Transkription könnte die komplexe Modifikation daher einen Einfluss haben.

Auch mit i⁶A, der Vorstufe von ms²i⁶A, wurden SILAC Pulldowns im selben Sequenzkontext durchgeführt. Hier konnten allerdings nahezu keine bindenden oder abgestoßenen Proteine identifiziert werden. In einem weiteren Experiment sollte hier noch geklärt werden ob die Ursache für die geringe Anzahl an bindenden Proteinen durch das im Experiment verwendete Blocking-Reagenz verursacht wurde. Bei diesem Reagenz, zur Reduktion unspezifischer Wechselwirkungen, handelt es sich um tRNA aus Hefe. Da sich in der tRNA aus Hefe jedoch auch natürliches i⁶A in relevanten Mengen findet, ist das Reagenz in diesem Fall nicht geeignet. Im Falle des ms²i⁶A ist die Verwendung der Konkurrenz-tRNA aus Hefe jedoch von Vorteil, da sich hier kein ms²i⁶A findet und durch die Anwesenheit von i⁶A im Konkurrenzsubstrat die Spezifität der ms²i⁶A- Binder noch erhöht wird. Im Laufe des Schreibens der Arbeit wurde der Pulldown von i⁶A haltiger sowie unmodifizierter tRNA mit DNA aus Spermien des Lachses als Konkurrenzsubstrat wiederholt. Auch hier konnten nahezu keine

spezifischen Binder identifiziert werden. Eine ganze Reihe von Proteinen wurde durch die Modifikation i^6A abgestoßen, es handelt sich jedoch nicht um Proteine die auch mit ms^2i^6A interagieren. Somit scheint die Spezifität der ms^2i^6A -Binder real zu sein. Die Identifizierten bindenden und abgestoßenen Proteine interagieren tatsächlich nur mit ms^2i^6A , nicht mit dessen Vorstufe i^6A .

Mit Hilfe einer *Gene-Ontology* Suche mit Panther konnte den von ms^2i^6A angezogenen und abgestoßenen Proteinen jeweils biologischen Funktionen zugeordnet werden. (Abb. 4.36)

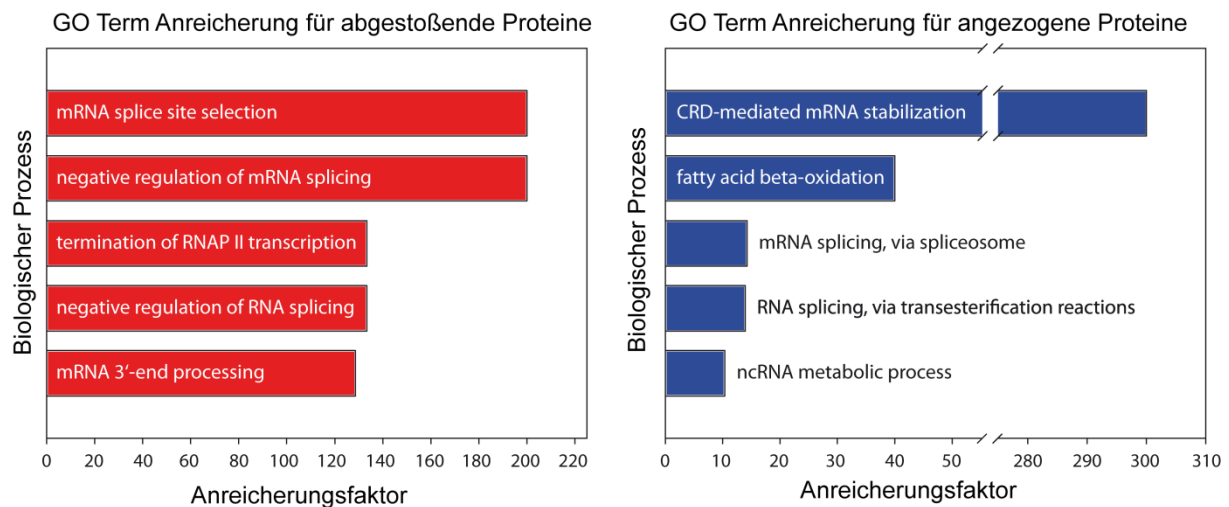


Abb. 4.36: Panther Cluster Analyse der biologischen Prozesse an denen ms^2i^6A bindende Proteine (blau) und durch ms^2i^6A abgestoßene Proteine (rot) beteiligt sind. Die Anreicherung des jeweiligen biologischen Prozesses ist als x-fache Anreicherung im Vergleich zu allen biologischen Prozessen dargestellt.

Unter den spezifischen Bindern von ms^2i^6A finden sich neben mRNA stabilisierenden Proteinen auch solche, die an der Fettsäure Betaoxidation beteiligt sind. Diese beiden biologischen Prozesse sind 300- bzw. 40-fach und damit sehr hoch gegenüber anderen biologischen Prozessen angereichert. Es ist daher sehr wahrscheinlich, dass einige ms^2i^6A - Binder an diesen Prozessen beteiligt sind. Unter den am höchsten angereicherten Bindern befinden sich zahlreiche Proteine die in verschiedene Arten des Spleißens involviert sind. Biologische Prozesse wie das Spleißen der mRNA über das Spleiceosom oder das RNA Spleißen über Transesterifizierungsreaktionen sind 10- bis 20-fach gegenüber anderen biologischen Prozessen angereichert. Unter den durch ms^2i^6A abgestoßenen Proteinen sind die biologischen Prozesse mRNA Spleißstellen Auswahl sowie die negative Regulation des mRNA Spleißens mit hoher Signifikanz

angereichert. Auch hier finden sich also zahlreiche Hinweise auf eine Beteiligung von ms^2i^6A bindenden und durch ms^2i^6A abgestoßenen Proteinen am Vorgang des Spleißens. Somit ergibt sich nicht nur durch einzelne Binder von ms^2i^6A , die mit dem Spleißen in Verbindung gebracht werden können, sondern auch durch die Gesamtheit aller Interaktoren ein deutlicher Hinweis auf weitere Funktionen von ms^2i^6A außerhalb der tRNA. Einen Einfluss auf das Spleißen bestimmter Proteine oder die Stabilisierung von mRNA sollte nur dann durch ms^2i^6A bindende Proteine beeinflussbar sein, wenn sich ms^2i^6A tatsächlich in mRNA befindet.

Um die Anreicherung spezifischer Proteine im SILAC Experiment zu validieren sollte anschließend der oben beschriebene Pulldown mit ms^2i^6A -enthaltender tRNA sowie mit unmodifizierter tRNA in unserer Gruppe wiederholt werden. Die gebundenen Proteine sollte anschließend nicht über LC-MS sondern über einen Western Blot identifiziert werden. Der Western Blot wurde mit kommerziell erhältlichen Antikörpern gegen die ms^2i^6A -Binderproteine HNRNPK und FAM120C, sowie gegen das abgestoßene Protein U2AF2, durchgeführt. Lediglich für den Antikörper gegen U2AF2 konnte ein Unterschied zwischen unmodifizierter tRNA und ms^2i^6A enthaltender tRNA erhalten werden. (Abb. 4.37)

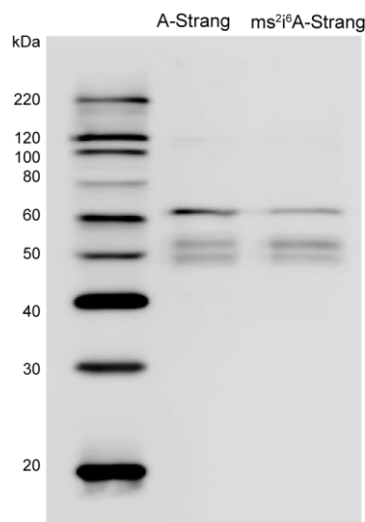


Abb. 4.37: Westernblot gegen U2AF2 nach einem Protein *Pulldown* mit unmodifizierter RNA und ms^2i^6A enthaltender RNA analog dem SILAC-Pulldown.

Man erkennt zwei unspezifische Banden bei ca. 50 kDa sowie eine spezifische Bande von U2AF2 bei etwas über 60 kDa. Die spezifische Bande ist bei der unmodifizierten RNA

etwas stärker ausgeprägt als bei der ms²i⁶A enthaltenden RNA. Mit Hilfe des Antikörpers konnte also im Pulldown mit unmodifizierter RNA mehr U2AF2 sichtbar gemacht werden, als im Pulldown mit ms²i⁶A enthaltender RNA.

Für die Binderproteine HNRNPK und FAM120C konnte wie schon erwähnt kein Unterscheid detektiert werden. Im Falle des HNRNPK könnte die Ursache in der Verwendung eines Antikörpers gegen die falsche Isoform von HNRNPK liegen. Von HNRNPK existieren drei Isoformen, der Antikörper jedoch bindet nur 2 davon. Im ursprünglichen SILAC-Pulldown konnte die tatsächlich bindende Isoform nicht klar identifiziert werden. Für die nicht durch den Antikörper nachweisbare Isoform konnten zwar keine spezifischen Peptide im SILAC Pulldown gefunden werden, da der Unterschied zwischen den Isoformen jedoch nur in wenigen Peptiden liegt, könnte es sich bei dem ms²i⁶A bindenden HNRNPK auch um die durch den Antikörper nicht nachweisbare Isoform handeln. Dies wäre eine Erklärung für die erfolglose Validierung von HNRNPK. Für FAM120C war das Signal bei einem vorherigen Test des Antikörpers gegen unbehandeltes Zelllysat bereits so schwach, dass durch den Pulldown möglicherweise nicht mehr genug Protein für ein solides Signal vorhanden war. Trotz mehrmaliger Versuche die Menge an angereichertem Protein im Pulldown zu steigern, konnte kein deutliches Signal für den FAM120C Antikörper erzielt werden. Somit konnte nur U2AF2 als von ms²i⁶A abgestoßenes Protein validiert werden.

Wie schon erwähnt konnten in einem SILAC Pulldown mit nuklearem Zelllysat nur wenige und nicht besonders signifikante Binder von ms²i⁶A identifiziert werden, während im Pulldown mit cytosolischen Lysat ein sehr umfassendes Interaktom erhalten wurde. Dies ist verwunderlich, da von den meisten Spleißingfaktoren eine Kernlokalisierung erwartet wird. Aus diesem Grund sollten einige der Binderproteine mit GFP fusioniert werden, um sie in Hela Zellen lokalisieren zu können. Nach dem Anfärben der Zellkerne mit DRAQ5 könnten das GFP Signal und das DRAQ5 Signal mit Hilfe eines Konfokalmikroskopes visualisiert werden (Abb. 4.38).

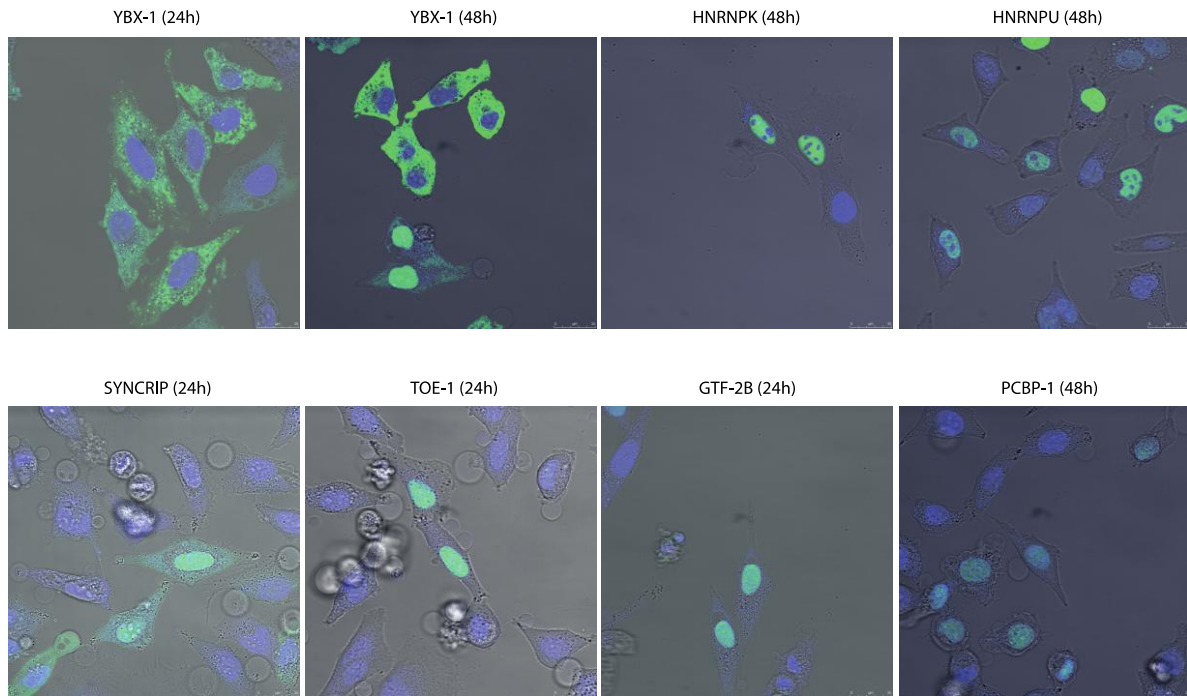


Abb. 4.38: Lokalisation einiger GFP-markierter, $ms^{2i6}A$ bindender Proteine zu den Zeitpunkten 24 h und 48 h nach der Transfektion von HeLa Zellen. Die Zellkerne wurden mit DAPI angefärbt und sind blau gefärbt. Die GFP-markierten Proteine sind grün angefärbt.

Man erkennt, dass das Protein YBX1 nach 24 h vor allem im Cytosol, nach 48 h jedoch teilweise auch im Zellkern lokalisiert ist. Somit ändert sich die Lokalisation abhängig vom Zeitpunkt der Visualisierung. Das Protein SYNCRIP ist nach 24 h sowohl im Zellkern als auch im Cytosol lokalisiert. Die Proteine HNRNPK, HNRNPU und PCBP1 konnten nur nach 48 h im Zellkern visualisiert werden. Auf Grund einer schlechten Transfektion oder verzögerter Reifung konnte nach 24 h noch kein deutliches GFP-Signal erhalten werden. Somit ist nicht auszuschließen, dass die Proteine nach 24 h auch im Cytosol lokalisiert gewesen sein könnten. Von HNRNPK ist bekannt, dass es sowohl im Nukleus als auch im Cytosol lokalisiert ist.^[184] Die Proteine TOE1 und GTF2B konnten nur im Zellkern lokalisiert werden. Somit sind viele Binderproteine kernlokalisiert, was verwunderlich ist, da sie lediglich im $ms^{2i6}A$ -Pulldown aus cytosolischem Lysat gefunden wurden. Es wäre möglich, dass viele Proteine sowohl im Cytosol als auch im Zellkern lokalisiert sind oder zu verschiedenen Zeitpunkten an verschiedenen Orten lokalisiert sind. Eine weitere Erklärung wäre, dass das cytosolische Zelllysat für den SILAC Pulldown leicht mit nuklearen Proteinen verunreinigt war. Möglicherweise binden diese Proteine im cytosolischen Lysat besser, da sie hier verdünnt vorliegen und dadurch die Bindung

möglicherweise erleichtert wird.

Wie zuvor schon beschrieben wird i^6A in Säugetieren mit Hilfe des Enzymes CDK5RAP1, SAM, Eisen III und Cystein zu ms^2i^6A thiomethyliert. Tomizawa *et al.* generierten einen *Knockout* von CDK5RAP1 in Mäusen um den Phänotyp charakterisieren zu können.^[149]

In Mäusen existieren zwei Computer-annotierte Spleißvarianten des CDK5RAP1. Es konnte gezeigt werden, dass eine Spleißvariante im Cytosol, die andere im Mitochondrium lokalisiert ist.^[147] Zudem wurde gezeigt, dass die cytosolische Variante, welcher am N-Terminus das mitochondrielle Importsignal fehlt, zwar dazu in der Lage ist ms^2i^6A zu produzieren, dass das Substrat i^6A im Cytosol aber fehlt.^[149] Ein *Knockout* der cytosolischen Variante würde folglich in Mäusen nicht zu reduziertem ms^2i^6A -Level führen. Im Menschen existieren 6 unterschiedliche Spleiß-Varianten des CDK5RAP1, eine davon ebenfalls mit cytosolischer Lokalisierung.^[149] Daher scheint es interessant auch in einer humanen Zelllinie die Auswirkung eines CDK5RAP1 *Knockouts* zu untersuchen. Hierzu wurden HAP1 Wildtyp Zellen sowie CDK5RAP1 *Knockout* HAP1 Zellen erworben und diese hinsichtlich ihrer Menge an i^6A und ms^2i^6A untersucht. (Abb. 4.39)

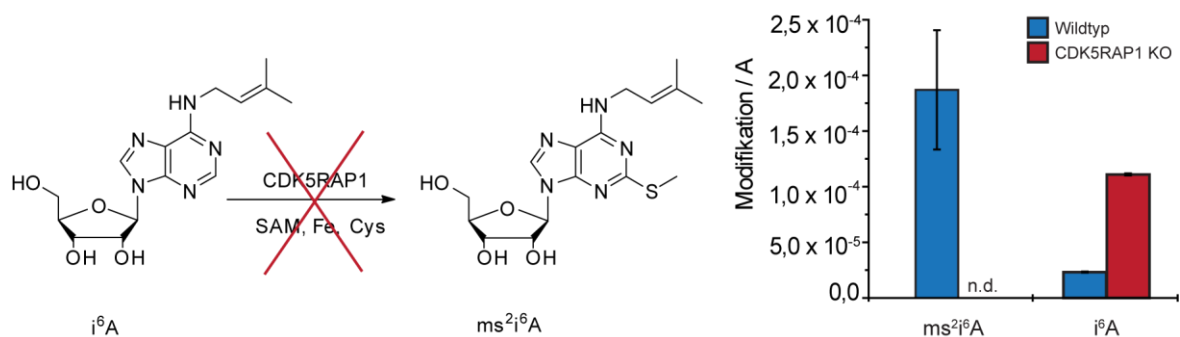


Abb. 4.39: Absolute Quantifizierung der Mengen an ms^2i^6A und i^6A jeweils in der Poly-A-RNA von HAP1 Wildtyp Zellen sowie CDK5RAP1 *Knockout* Zellen.

Man erkennt, dass nach dem *Knockout* von CDK5RAP1 in den Hap-Zellen kein ms^2i^6A mehr detektierbar ist. Die Mengen an i^6A dagegen nehmen in der *Knockout*-Zelllinie deutlich zu. Somit bleibt im CDK5RAP1 KO ein Großteil des i^6A bestehen, welches in Anwesenheit von CDK5RAP1 zu ms^2i^6A umgesetzt worden wäre.

Mit Hilfe dieser Daten konnte der *Knockout* als erfolgreich gewertet werden und weitere Untersuchungen zu Auswirkungen eines Mangels an CDK5RAP1 und damit an ms^2i^6A

unternommen werden.

Im Folgenden wurde eine proteomweite Analyse der HAP1 Wildtyp sowie der HAP1 CDK5RAP1 KO Zellen vorgenommen. Hierzu wurden die Zellen lysiert, die Disulfidbrücken der Proteine reduziert und alkyliert und tryptisch verdaut und anschließend über LC-MS/MS analysiert. Durch eine Fraktionierung der Zelllysate in cytosolisches, mitochondrielles und nukleäres Zelllysate, sowie über eine Aufreinigung der Proben vor der LC-MS-Analyse durch STAGE Tips und eine stufenweise Elution konnte die Abdeckung der gemessenen Proteine von 15% auf etwa 40% gesteigert werden. Die Quantifizierung der Proteine erfolgte über markierungsfreie Quantifizierung (LFQ) mit drei biologischen Replikaten. Durch den Vergleich von HAP1-Wildtyp Zellen mit der *Knockout* Zelllinie konnten zahlreiche differentiell exprimierte Proteine mit hoher Signifikanz identifiziert werden. Die An- oder Abwesenheit von CDK5RAP1 scheint daher einen großen Einfluss auf die Biosynthese vieler Proteine zu haben. Mit Hilfe der zu jedem identifizierten Protein zugeordneten GO-Terme konnten zudem bestimmte Gruppe an Proteinen, die im Wildtyp angereichert wurden, identifiziert werden. (Abb. 4.40)

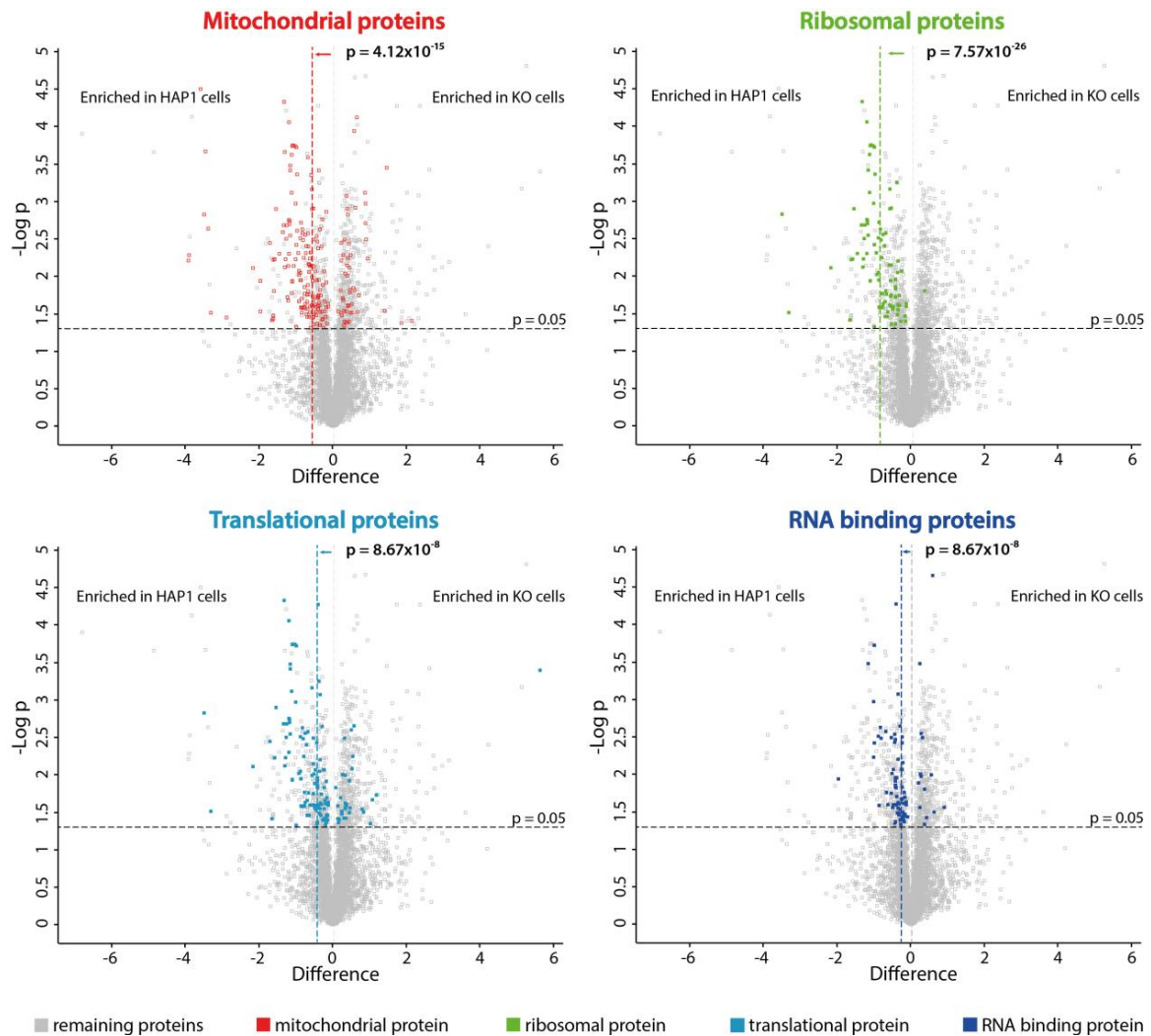


Abb. 4.40: Plot der in HAP1 Wildtyp-Zellen und CDK5RAP1 KO Zellen spezifisch angereicherten Proteine. Je größer der Unterschied zum Ursprung 0 der x-Achse und je größer der $-\log p$ desto stärker würde das Proteine in den Wildtyp- bzw. *Knockout*-Zellen angereichert. Farblich markiert sind jeweils die im Wildtyp angereicherten Proteine einer bestimmten biologischen Funktion bzw. einer Klasse von Proteinen. Der Grad der Anreicherung ist durch den Wert p angegeben.

Mit einer sehr hohen Signifikanz von $p = 7.57 \times 10^{-26}$ konnten ribosomale Proteine eher im CDK5RAP1 Wildtyp Zellen als im Knockout gefunden werden. Auch mitochondriale Proteine sind mit einer relativ hohen Signifikanz $p = 4.12 \times 10^{-15}$ in den Knockoutzellen abgereichert. Dies spricht dafür, dass ein Fehlen von CDK5RAP1 und damit von ms^2i^6A zu einer Veränderung in der Synthese von ribosomalen und mitochondrialen Proteinen führt. Im Falle der mitochondrialen Proteine deckt sich dies mit der Entdeckung von Wei *et al.* welche eine Fehlfunktion der mitochondrialen Translation bedingt durch ein Fehlen von ms^2i^6A vermuten.^[149]

Für translationale Proteine und RNA bindende Proteine erkennt man einen ähnlichen, wenn auch nicht ganz so stark ausgeprägten Trend. Auch hier scheint das Fehlen von CDK5RAP1 und damit von ms²i⁶A den Syntheseweg der zugehörigen Proteine zu stören. Sowohl für die im CDK5RAP1 KO an- als auch für die im KO abgereicherten Proteine wurde zudem eine funktionelle Cluster Anreicherungsanalyse mit DAVID vorgenommen. Hierzu wurden alle signifikant an- oder abgereicherten Proteine berücksichtigt. (Abb. 4.41)

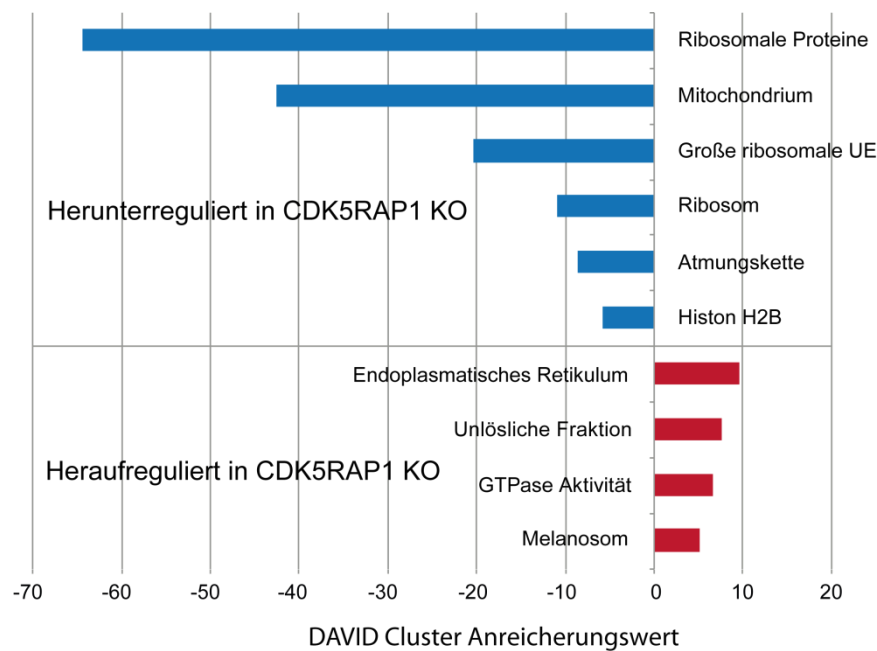


Abb. 4.41: *Gene-Ontology* Analyse der im HAP1 Wildtyp (blau) und im CDK5RAP1 KO (rot) angereicherten Proteine. Je größer der Cluster Anreicherungs-wert, desto mehr Proteine dieser Proteinklasse konnten identifiziert werden.

Auch hier erkennt man deutlich eine Abreicherung von ribosomalen und mitochondrialen Proteinen bei einem Mangel an CDK5RAP1. Auch Proteine der Atmungskette und Histon H2B assoziierte Proteine tauchen im CDK5RAP1 KO in deutlich geringeren Mengen auf. Durch die *Gene-Ontology* Analyse erfährt man auch etwas über Proteine, die im *Knockout* vermehrt produziert werden. So scheint ein *Knockout* von CDK5RAP1 einen positiven Einfluss auf Proteine des endoplasmatischen Retikulums, solche mit GTPase Aktivität und Proteine des Melanosomes zu haben. Die Anreicherungs-werte sind jedoch deutlich schwächer ausgeprägt.

Im weiteren Verlauf des Projektes wurde zur Charakterisierung des Phänotyps auch

eine RNA-Sequenzierung von RNA Fragmenten aus HAP Wildtyp sowie CDK5RAP1 KO Zellen vorgenommen. Hierzu wurde zunächst eine *Paired-end* Bibliothek mit RNA aus je drei unterschiedlichen Hap Wildtyp und Hap CDK5RAP1 KO Zelllinien generiert und die durch RNA-Sequenzierung erzeugten Daten anschließend mit Hilfe der Programme Tophat2 mit den hg19 Referenzgenom abgeglichen und die relative Expression mit Cufflinks quantifiziert. Anschließend wurde in den Proteomdaten nach Proteinen gesucht, welche sowohl in der Proteom- als auch in der Transkriptomanalyse jeweils im Wildtyp oder *Knockout* angereichert sind. Interessanterweise konnte ein großer Überlapp zwischen den an- bzw. abgereicherten Proteinen in beiden Datensätzen erzielt werden. (Abb. 4.42)

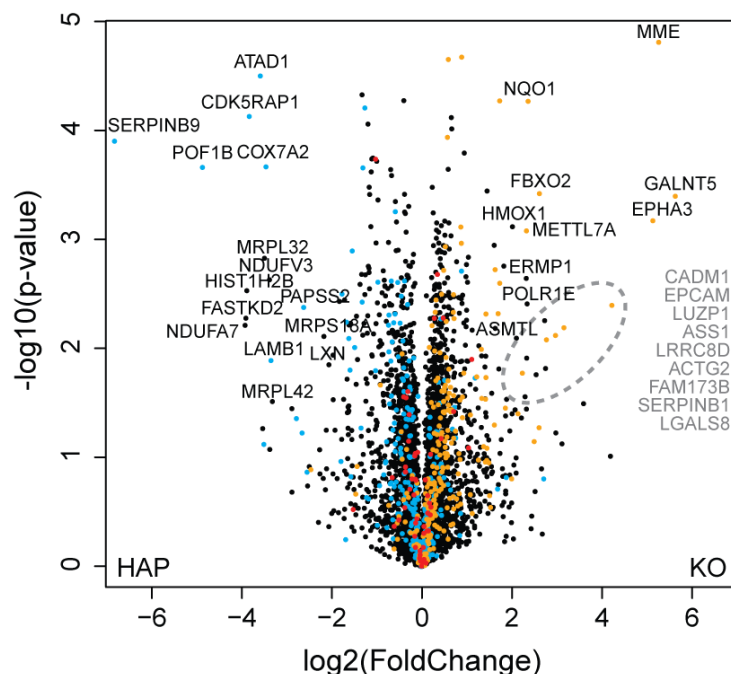


Abb. 4.42: Vulkan-Plot der in der Proteomanalyse in HAP1 Wildtyp-Zellen und CDK5RAP1 KO Zellen spezifisch angereicherten Proteine. Je größer der Unterschied zum Ursprung 0 der x-Achse und je größer der $-\log p$ desto stärker würde das Proteine in den Wildtyp- bzw. *Knockout*-Zellen angereichert. Blau markiert sind Proteine die sowohl in der Proteom- als auch in der Transkriptomanalyse in den Wildtypzellen signifikant angereichert wurde. Proteine die in beiden Datensätzen im CDK5RAP1 KO signifikant angereichert werden konnten sind orange markiert.

Zahlreiche Proteine, die bei der Proteomanalyse spezifisch im Wildtyp oder Knockout angereichert werden konnten, finden sich auch auf der Ebene des Transkriptom wieder. Somit wurde nicht nur die Proteinbiosynthese an sich gestört sondern die Veränderung durch den Knockout ist schon auf der Ebene des Transkriptes bemerkbar.

Dieses Ergebnis kann zudem auch als Validierung der Proteom- sowie der Transkriptomanalyse verwendet werden. Da beide Datensätze ähnliche Ergebnisse liefern, scheinen die Experimente funktioniert zu haben.

Mit Hilfe einer funktionellen Cluster Anreicherungsanalyse mit DAVID konnten die für den CDK5RAP1 *Knockout* an- oder abgereicherten Transkripte in Proteingruppen zusammengefasst werden. (Abb. 4.43)

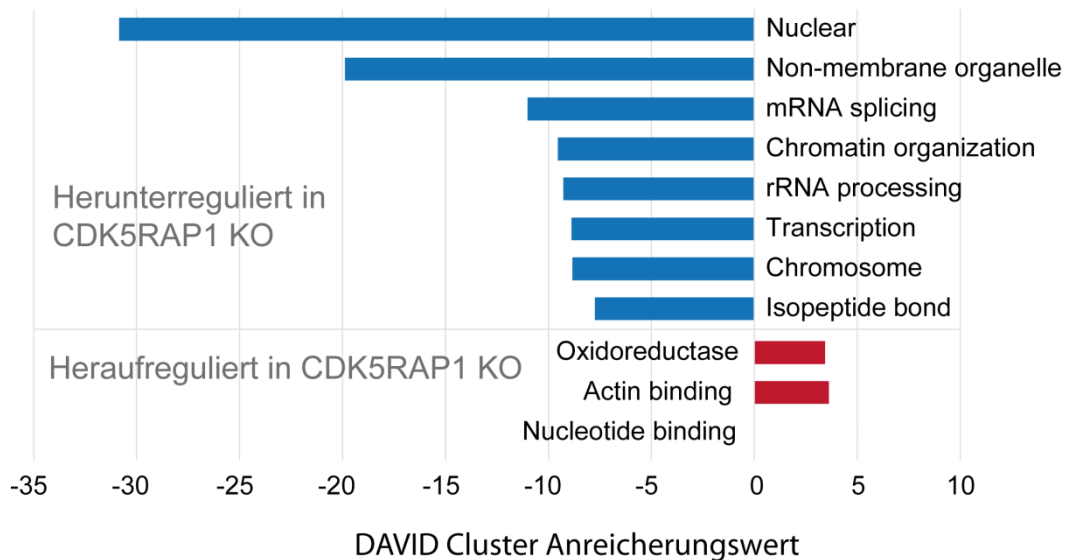


Abb. 4.43: Cluster Anreicherungsanalyse der in Hap Wildtyp (blau) und im CDK5RAP1 KO (rot) angereicherten Transkripte. Je größer der Cluster Anreicherungswert, desto mehr Transkripte dieser Proteinklasse konnten identifiziert werden.

Man erkennt, dass nukleare Faktoren, in das Spleißen von mRNA involvierte Proteine, sowie auch rRNA prozessierende und an der Transkription beteiligte Proteine im Wildtyp angereichert sind. Auch gespleißte Proteine selbst sind im CDK5RAP1 KO abgereichert und damit im Wildtyp angereichert. Interessanterweise ergibt sich hier eine ähnliche Aussage, wie bei der ursprünglichen GO-Analyse der Binderproteine von ms²i⁶A und auch wie für die Cluster Anreicherungsanalyse der an- und abgereicherten Proteine der Proteomanalyse des CDK5RAP1 *Knockouts*.

Sowohl die Analyse der Transkripte der CDK5RAP1 Wildtyp und der *Knockout* Zelllinie, als auch die Analyse des Proteomes konnten zudem Hinweise auf differenziell gespleißte Proteine geben. Für einige Proteine konnten sowohl auf Transkriptebeine als auch auf der Ebene der translatierten Proteine, die eine differenzielle Regulation festgestellt

werden. Im weiteren Verlauf dieses Projektes wurde daher in den RNA-Sequenzierungsdaten gezielt nach alternativen Spleißmustern gesucht. Neben 91 anderen differenziell gespleißten Proteinen konnte alternatives Spleißen bei dem mitochondrialen Transkriptionsfaktor A (TFAM) festgestellt werden. Hier ergibt sich ein weiterer Hinweis auf eine mitochondriale Funktion von ms^2i^6A . Zudem konnte TFAM schon im SILAC-Pulldown als ms^2i^6A Binder identifiziert werden. Somit könnte ms^2i^6A hier eine Vermittlerrolle einnehmen. Durch dessen Anwesenheit könnten spezielle Formen des Spleißens getriggert und dadurch ein bestimmtes Proteine alternativ gespleißt werden. Im Falle des TFAM würde sich also die Anwesenheit von ms^2i^6A indirekt auf die Synthese mitochondrialer Proteine auswirken.

Im weiteren Verlauf des Projektes sollte auch der Frage nach der Lokalisierung der Modifikation ms^2i^6A in mRNA nachgegangen werden. Eine Lokalisierung von ms^2i^6A in kodierenden Bereichen scheint auf Grund der starken Effekte der Modifikation auf die Basenpaarung und die Ausbildung und Stabilisierung von Strukturen, welche aus tRNA bekannt sind, als eher unwahrscheinlich. Viel mehr könnte es in Introns lokalisiert sein. Aus diesem Grund wurde ein Experiment mit Isoginkgetin durchgeführt. Es handelt sich um einem zellpermeablen Inhibitor des pre-mRNA Spleißens, der Spleißosom vermittelte Prozesse am pre-Spleißosom unterbindet. Wird der Inhibitor zu den Zellen gegeben, kann also kein pre-mRNA Spleißen mehr stattfinden. Dadurch bleiben die Introns in der pre-mRNA (Abb. 4.44 A). Wäre ms^2i^6A also in den Introns lokalisiert, sollte nach einer Behandlung mit dem Inhibitor vermehrt ms^2i^6A in der pre-mRNA gefunden werden. Zunächst wurde Isoginkgetin auf seine Wirksamkeit getestet. Das Unterbinden des Spleißvorgangs wurde über die Messung der Verhältnisse von mRNA und pre-mRNA an einem Beispiel-Gen überprüft. Schon bei einer Isoginkgetin Konzentration von 50 μM konnte nach 6 h ein etwa 1.5-facher Anstieg der pre-mRNA im Vergleich zur DMSO-Kontrolle beobachtet werden. Höhere Konzentrationen an Isoginkgetin waren aufgrund der Sensitivität der Zellen gegenüber DMSO nicht möglich. Bei höheren Konzentrationen sterben die Hap Zellen zunehmend. Daher wurden die Hap Zellen mit 50 μM Isoginkgetin behandelt und von diesen Zellen wie auch von den unbehandelten Hap Zellen die Poly-A-RNA isoliert. Anschließend wurden die Mengen an ms^2i^6A sowie auch an i^6A in den mit Isoginkgetin behandelten und in den unbehandelten Zellen bestimmt. (Abb. 4.44 B)

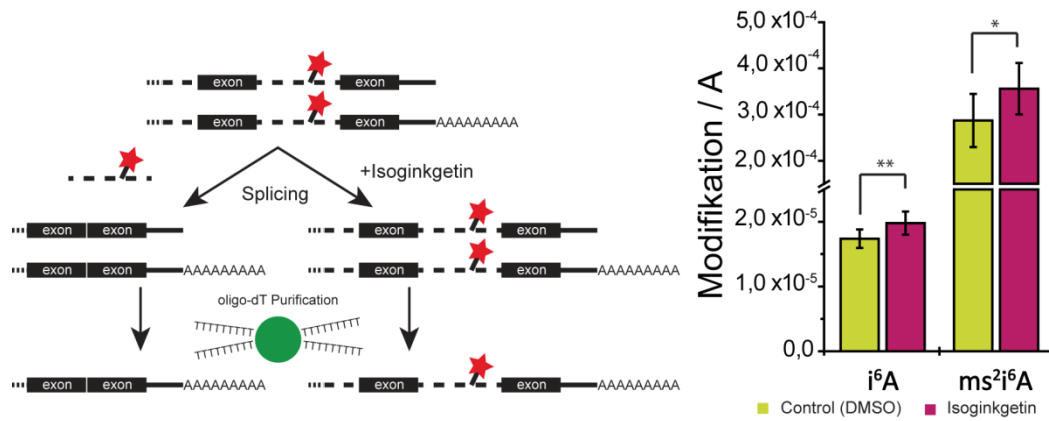


Abb. 4.44: A) Darstellung der Wirkungsweise von Isoginkgetin auf das Spleißen von pre-mRNA. B) Quantifizierung der Mengen an i^6A und ms^2i^6A in mit Isoginkgetin behandelten Hap Zellen sowie in unbehandelten Hap Zellen. Die Zunahme der Mengen nach Behandlung mit dem Inhibitor ist in beiden Fällen (i^6A ($p=0.0095$) und ms^2i^6A ($p=0.022$)) signifikant.

Man erkennt sowohl für ms^2i^6A als auch für i^6A eine Zunahme der Modifikation in pre-mRNA nach einer Behandlung der Hap Zellen mit Isoginkgetin. Die Zunahme ist im Vergleich zu unbehandelten Zellen in beiden Fällen signifikant. Daraus kann man ableiten, dass sich zumindest ein Teil des ms^2i^6A wie auch des i^6A in Introns oder *skipped* Exons befinden muss und ohne den Spleiß-Inhibitor normalerweise durch das Spleißen entfernt wird.

Im Verlauf des Projektes konnten zahlreiche Hinweise auf eine Funktion der Modifikation ms^2i^6A außerhalb der tRNA gefunden werden. Durch einen SILAC-Pulldown mit ms^2i^6A enthaltender tRNA konnten viele ins Spleißen involvierte Proteine signifikant an-oder abgereichert werden. Auch eine GO-Analyse aller Binder ergab eine starke Anreicherung von Proteinen, die in verschiedenste Arten des Spleißens involviert sind. Unter den abgestoßenen Proteinen fanden sich zahlreiche Proteine die an der negativen Regulation des Spleißens beteiligt sind. Einige Binderproteine wurden in GFP-Plasmide kloniert und nach einer Transfektion in HEK-Zellen am Konfokalmikroskop visualisiert. Abhängig vom Zeitpunkt konnten die Binder sowohl im Zellkern als auch im Cytosol lokalisiert werden. Mit Hilfe eines *Knockouts* von CDK5RAP1 in humanen HAP Zellen konnte ein Phänotyp ohne ms^2i^6A erhalten werden. Mit Hilfe einer Analyse des Proteoms der *Knockout*- wie auch der Wildtypzelllinie konnten Proteingruppen identifiziert werden, die in Abwesenheit von CDK5RAP1 und damit von ms^2i^6A stark

abgereichert sind. Die Abwesenheit von CDK5RAP1 hat eine starke Auswirkung auf die Synthese von mitochondrialen, ribosomalen, translationalen und RNA bindenden Proteinen. Eine RNA Sequenzierung des Transkriptoms von CDK5RAP1 KO sowie aus den Wildtyp Zellen ergab ein sehr ähnliches Bild. Zahlreiche Proteine die bei der Proteomanalyse im *Knockout* abgereichert waren, waren auch auf der Transkriptebene weniger abundant. Eine Analyse der im *Knockout* abgereicherten Transkripte ergab zudem ein ähnliches Bild wie die ursprüngliche Analyse der SILAC-Daten und lieferte erneut Hinweise auf einen Einfluss der Modifikation auf den Vorgang des Spleißens. Durch weitere Analysen der Sequenzierungsdaten konnten zudem alternativ gespleißte Proteine entdeckt werden, darunter ein mitochondrialer Transkriptionsfaktor, der schon im SILAC Plot unter den Bindern zu finden war. Die Anwesenheit von ms^2i^6A scheint sich somit indirekt auf die Synthese mitochondrialer Proteine auswirken. Durch die Modifikation könnten spezielle Formen des Spleißens getriggert und dadurch bestimmte Proteine alternativ gespleißt werden. Mit Hilfe eines des Spleiß-Inhibitors Isoginkgetin konnten zudem Hinweise zur Lokalisierung ms^2i^6A in der mRNA gewonnen werden. Die Modifikation scheint sich in Introns oder *skipped* Exons zu befinden, da es durch ein Unterbinden des Spleißvorgangs zu einer Anreicherung von ms^2i^6A in pre-mRNA kommt. Somit scheint sich ms^2i^6A in den Introns der pre-mRNA zu befinden und das alternative Spleißen spezieller Proteine zu beeinflussen.

4.4.4 Zusammenfassung und Projektbeiträge

Die Synthese der benötigten RNA-Sequenzen wurde von Simon Geiger durchgeführt. Die *SILAC-Pulldown* Experimente zur Identifizierung von ms^2i^6A bindenden Proteinen wurden von Raghu Edupuganti aus der Gruppe von Michiel Vermeulen vorgenommen.

Die GO-Analysen, Lokalisationsexperimente der bindenden Proteine, sowie die Charakterisierung der CDKRAP1 KO Zellen über einer Analyse des Proteomes wurden in enger Zusammenarbeit mit Simon Geiger vorgenommen. Die Validierung des *SILAC Pulldowns* über Western Blot erfolgte in Zusammenarbeit mit Simon Geiger, Matthias Kurz und Martin Rossa.

Die Quantifizierung der RNA Modifikationen wurden von Katharina Iwan und Sarah Schiffers durchgeführt. Die Analyse des Transkriptoms der CDK5RAP1 KO Zelllinie erfolgte durch Markus Müller und Olga Kotljarova. Die weitere Charakterisierung der KO Zelllinie sowie die Versuche zum alternativen Spleißvorgang wurden von Angie Kirchner und Katharina Iwan durchgeführt.

4.4.5 Material und Methoden

Medien, Lösungen und Puffer

Bezeichnung	Zusammensetzung
DMEM Medium	DMEM - <i>high glucose</i> (4500 mg/L Glukose) Fertigmedium (<i>Sigma Aldrich</i>), 56 mL FBS (<i>Gibco</i>), 10.25 mL L-Alanyl-L-Glutamin (200 mM, <i>Sigma-Aldrich</i>) 5.6 mL Antibiotic-Antimycotic (<i>Sigma-Aldrich</i>) oder 5.6 mL Penicillin-Streptomycin (<i>Sigma Aldrich</i>)
RPMI Medium	RPMI-1640 Medium, 56 mL FBS (<i>Gibco</i>), 10.25 mL L-Alanyl-L-Glutamin (200 mM, <i>Sigma-Aldrich</i>) 5.6 mL Antibiotic-Antimycotic (<i>Sigma-Aldrich</i>) oder 5.6 mL Penicillin-Streptomycin (<i>Sigma Aldrich</i>)
DPBS	Dulbecco's PBS (<i>Sigma Aldrich</i>)
LB Medium	1.0% (w/v) Pepton, 0.5% (w/v) Hefeextrakt, 1.0 % (w/v) NaCl; pH = 7.5; nach dem Autoklavieren 0.1% Antibiotika-Stocklösung
LB Agar	1.5% (w/v) Agar-Agar in LB-Medium
SOC Medium	0.5% (w/v) Pepton, 0.5% (w/v) Hefeextrakt, 10 mM NaCl, 2.5 mM KCl; nach dem Autoklavieren 5.0 mM MgCl ₂ , 5.0 mM MgSO ₄
EMSA-Lysispuffer	10 mM Tris-HCl (pH 7.5), 150 mM KCl, 0.5 mM EDTA, 0.5% (v/v) NP-40, 1 × Roche Complete Protease Inhibitor, 1% (v/v) Phosphatase Inhibitor Cocktail 2 und 3 (<i>Sigma Aldrich</i>)
VM-Puffer	10 mM Tris-HCl (pH 7.5), 150 mM KCl, 1.5 mM MgCl ₂ , 0.1 % (v/v) NP-40, 0.5 mM DTT, 1 × Roche Complete Protease Inhibitor, 1% (v/v) Phosphatase Inhibitor Cocktail 2 und 3 (<i>Sigma Aldrich</i>), 0.4 u/μLRNasin (nur in Inkubationsschritten)
Pulldown-Lysispuffer	10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5 mM EDTA, 0.5% (v/v) NP-40, 1 × Roche Complete Protease Inhibitor, 1% (v/v) Phosphatase Inhibitor Cocktail 2 und 3 (<i>Sigma Aldrich</i>)

Dilution Buffer	10 mM Tris·HCl (pH 7.5), 150 mM NaCl, 0.5 mM EDTA
Proteinase K Puffer	10 mM Tris·HCl (pH 8.0), 1.0 mM EDTA, 0.5% (w/v) SDS
Proteinase K Stock	50% (v/v) Glycerin, 20 mM Tris·HCl (pH 8.0), 1 mM CaCl ₂ , 2 mg/mL Proteinase K (<i>GeneON</i>)
10 × TBE	890 mM Tris·HCl (pH 8.0), 890 mM NaH ₂ BO ₃ , 20 mM EDTA
2 × RNA Ladepuffer	2 × TBE (178 mM Tris·HCl (pH 8.0), 178 mM NaH ₂ BO ₃ , 4 mM EDTA), 12% (w/v) Ficoll 400, 7 M Harnstoff

Antibiotika

Antibiotika wurden als 1000 × Stocklösungen vorbereitet und bei -20 °C gelagert.

Carbenicillin:

100 mg/mL in H₂O

Chloramphenicol:

34 mg/mL in EtOH_{abs}

Kanamycin:

50 mg/mL in H₂O

Spectinomycin:

100 mg/ml in H₂O

Zellkultur HeLa Zellen

Die humane Zelllinie HeLa wurden bei -80 °C in einem DMSO-haltigen Medium unter flüssigem Stickstoff gelagert. Unter Standardbedingungen wurde die Zelllinien bei 37 °C und 5% CO₂ in einem Inkubator bis zu 80% Konfluenz kultiviert. Als Medium für die Zellen diente ein DMEM-Fertigmedium (Invitrogen 41966-029), dem 10% (v/v) fötales Kälberserum und 1% (v/v) Penicillin-Streptomycin hinzugefügt wurden. Für die Abspaltung der Zellen wurde 25% Trypsin-EDTA Lösung (Gibco) und zum Waschen steriler DPBS-Puffer (Sigma) verwendet. Die Zellen wurden standardmäßig 5 min mit 5 mL Trypsin bei 37°C abgespalten. Die Suspension wurde 10 min bei 300 xg und RT zentrifugiert und das Medium entfernt. Das Zellpellet wurde im entsprechenden Medium resuspendiert und ein Splitting von 1/10 durchgeführt oder das Zellpellet für Zelllysate verwendet.

HeLa Zelllysat

Die HeLa Zellen wurden nach dem Ernten mit 2 mL DPBS gewaschen und anschließend mit je 150 µL Lysis Puffer (50 mM Tris-HCl pH8, 150 mM NaCl, 1% NP40 und Proteaseinhibitor +/- EDTA (Roche) auf Eis lysiert. Nach 30 min wurden die Zellen mit Hilfe eines Schabers vom Boden des Kultivierungsgefäßes abgeschabt und in Reaktionsgefäße überführt. Anschließend wurde 30 min bei 4 °C zentrifugiert und das Lysat zu klären.

Pulldown Experiment für den Western Blot

Für den RNA pulldown wurden jeweils 5 mg biotinylierte RNA an 50 mL Streptavidin Sepharose *high performance beads* (GE Healthcare) immobilisiert und diese in einem Gesamtvolumen von 1.5 mL RNA Bindungspuffer (150 mM NaCl, 50 mM Hepes-HCl pH 7.5, 0.1% NP40 (v/v), 10mM MgCl₂) für 1 h bei 4°C inkubiert. Die Streptavidin Sepharose *high performance beads* wurden zuvor zweimal mit RNA Bindungspuffer (150 mM NaCl, 50 mM Hepes-HCl pH 7.5, 0.5% NP40 (v/v), 10mM MgCl₂) gewaschen und mit 500 mL RNA Bindungspuffer and RNase Inhibitor (0.4 U) inkubiert. Nach der Immobilisierung der RNA auf den Partikeln wurde einmal mit RNA Waschpuffer (250 mM NaCl, 50 mM Hepes-HCl pH 7.5, 0.5% NP40 und 10 mM MgCl₂) und zweimal mit Protein Inkubationspuffer (150 mM KCl, 1.5 mM MgCl₂, 0.1% (v/v) NP-40, 0.5 mM DTT, 10 mM Tris-HCl pH 7.5 und Proteaseinhibitor +/- EDTA (Roche)) gewaschen. Anschließend wurden die Partikel mit der immobilisierten RNA mit 1.04 mg HeLa Lysat in in einem Volumen von 1500 µl Protein Bindungspuffer (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 und Proteaseinhibitor (Roche)) in Anwesenheit von 15 µL RNasin (0.4 U) bei Raumtemperatur für 30 min und bei 4°C für 90 min inkubiert. Daraufhin wurde dreimal mit je 1 mL Protein Bindungspuffer gewaschen und die *beads* zweimal mit 12 µL 1x SDS-loading Puffer für 3 min bei 65°C abgekocht. Anschließend wurden die Proteine in einem SDS Gel für 5 min bei 180 V und 45 min bei 140 V aufgetrennt.

Western Blot

Die Analyse über Western Blot erfolgte mit Hilfe eines Wet-Blots auf eine Nitrozellulosemembran. Dabei wurde für 16 h bei 120 V in Wet-Blot Puffer (25 mM Tris pH 8.3, 192 mM Glycin, 0.037% % (v/v) SDS, 10% Methanol) übertragen. Anschließend wurde mit 5% Milchpulver in TBS-T (0.5% Tween 20) für 1 h bei Rt geblockt. Anschließend wurden die Primärantikörper gegen HNRNPK (Thermo Fischer, PA5-27522), gegen FAM120c (Sigma, SAB2107574-100UL) und gegen U2AF2 (Sigma, SAB2701232-100UL) im vom Hersteller angegebenen Verdünnungsverhältnis in TBS-T Puffer zugegeben und bei 4°C über Nacht inkubiert. Nachdem dreimal mit TBS-T gewaschen wurde, wurde der sekundäre Antikörper Anti Rabbit A6667 (Sigma) in einer Verdünnung von 1:10000 in TBS-T zugegeben und 1 h bei Rt inkubiert. Nachdem wiederum dreimal mit TBS-T gewaschen wurde wurde mit Hilfe des ECL-Kits (GE) und einem LAS-3000 *Image Reader* visualisiert.

Bradford-Assay

Der Bradford-Assay dient der Bestimmung der Konzentration von Zelllysaten oder Proteinpräparationen und basiert auf der Verschiebung des Absorptionsmaximums von Coomassie-Brilliant-Blau von 465 nm auf 595 nm durch die Anlagerung von anionischen Gruppen des Proteins an den Farbstoff. Es wurden 950 µL Bradford-Reagenz mit 50µL Proteinlösung versetzt, bei Raumtemperatur für 5 min inkubiert und die Absorption bei 595 nm in einem Spektrophotometer bestimmt. Für die Konzentrationsbestimmung des Proteins wurde eine Eichgerade basierend auf einer BSA-Verdünnungsreihe gemessen und für die Berechnung der Protein Konzentration verwendet.

Klonierung der Binderproteine

Plasmide

Plasmide mit Inserts zur Expression von HNRNPK, YBX1, SYNCRIP, GTF2B, HNRNPU, PCBP1 und TOE1 wurden von *DNASU* (Tempe, AZ) bezogen und sind in der nachfolgenden Tabelle aufgelistet.

Tabelle 4.4: Übersicht der Plasmide für die Klonierung der Binderproteine.

Insert	Seriennummer	Vektor	Resistenz	Quelle
HNRNPK	HsCD00622816	pENTR223	Spectinomycin	DNASU, Tempe, AZ
YBX1	HsCD00040576	pDONR221	Kanamycin	DNASU, Tempe, AZ
SYNCRIP	HsCD00513121	pENTR223	Spectinomycin	DNASU, Tempe, AZ
GTF2B	HsCD00509797	pENTR223	Spectinomycin	DNASU, Tempe, AZ
PCBP1	HsCD00288587	pENTR223	Spectinomycin	DNASU, Tempe, AZ
TOE1	HsCD00399441	pDONR223	Spectinomycin	DNASU, Tempe, AZ
HNRNPU	HsCD00603898	pANT7_cGST	Carbenicillin	DNASU, Tempe, AZ
-	-	pDONR201	Kanamycin	Life Technologies
-	12281-010	pDEST47	Carbenicillin	Life Technologies
HNRNPK	-	pDEST007	Carbenicillin	

Transformation von chemisch kompetenten *E. coli* Zellen

Chemisch kompetente Zellen (50 μ L) wurden aus der Lagerung (-80 °C) entnommen und auf Eis ca. 5 min aufgetaut. Zu den Zellen wurden 50 bis 300 ng Plasmid-DNA bzw. 1 μ L Klonase-Reaktion gegeben und die Mischung 30 min auf Eis inkubiert. Nach einem Hitzeschock (42 °C, 30 s) wurden die Zellen wieder auf Eis gekühlt und mit SOC Medium (450 μ L) versetzt, danach bei 37 °C für 1 h inkubiert und 50 bis 150 μ L der Reaktion auf LB-Agarplatten ausgestrichen. Die Platten wurden über Nacht (ca. 16 h) bei 37 °C inkubiert.

Bakterien-Übernachtskultur

Für Bakterien-Übernachtskulturen wurden 4 bis 6 mL LB-Medium in ein 14 mL-Kulturröhrchen (*Greiner bio-one*) oder 50 mL LB-Medium in einen 250 mL Erlenmeyerkolben gegeben und mit dem zur Selektion nötigen Antibiotikum versetzt. Mit einer sterilen Pipettenspitze wurde eine einzelne Kolonie selektiert und die Flüssigkultur damit inokuliert. Die Kultur wurde bei 37 °C und 200 rpm für 14 bis 16 h inkubiert.

Gateway-Klonierung

Das *Gateway Cloning* System basiert auf homologer Rekombination zwischen zwei Vektoren. Es ist abgeleitet von den Rekombinationsreaktionen, die bei der Insertion der Phagen DNA des Bakteriophagen λ in das Genom des befallenen Wirtes stattfinden.^[191, 192] Es findet hierbei ein Austausch an definierten Sequenzabschnitten des bakteriellen Chromosoms (*attB*) und der Phagen DNA (*attP*) statt, die über einen sieben Nukleotide langen homologen Abschnitt verfügen. Der natürliche Prozess erzeugt zwei hybride Sequenzen „links“ (*attL*) und „rechts“ (*attR*) des insertierten Abschnittes.

Unter Verwendung von Plasmiden, die über die nötigen orthogonalen *att sites* verfügen, lassen sich in LR-, bzw. BP-Reaktionen schnell und zuverlässig Expressionsvektoren erzeugen (Abb. 4.45). Die verwendeten *Destination Vectors* ermöglichen durch ihre *ccdB* (*control of cell death*) Kassette eine Selektion der wie gewünscht klonierten Plasmide, da nur im Falle einer erfolgreichen LR Reaktion ein Plasmid entsteht, das über das Resistenzgen des *Destination* Plasmids verfügt, ohne gleichzeitig das *ccdB* Toxin zu codieren.

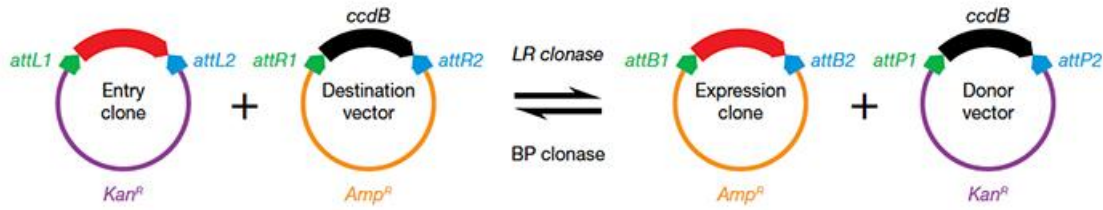


Abbildung 4.45: Schematische Darstellung des Gateway Cloning Systems.

Ziel der Gateway Klonierung war es, alle Inserts (*HNRNPK*, *YBX1*, *SYNCRIP*, *GTF2B*, *HNRNPU*, *PCBP1* und *TOE1*) mit einem C-terminalen GFP Tag zu versehen und in einen für Säugerzellen geeigneten Expressionsvektor zu klonieren. Aufgrund dieser Vorgaben wurde der Vektor pDEST47 (*Thermo Scientific*) als Expressionsplasmid ausgewählt. *HNRNPU* wurde zunächst in einer BP-Reaktion aus dem erhältlichen Vektor (pANT7_cGST), der lediglich über *attB sites* verfügt, in einen geeigneten intermediären Vektor (pDONR201) kloniert. Darauf folgend konnten alle Gene in LR-Reaktionen in den Zielvektor pDEST47 kloniert werden.

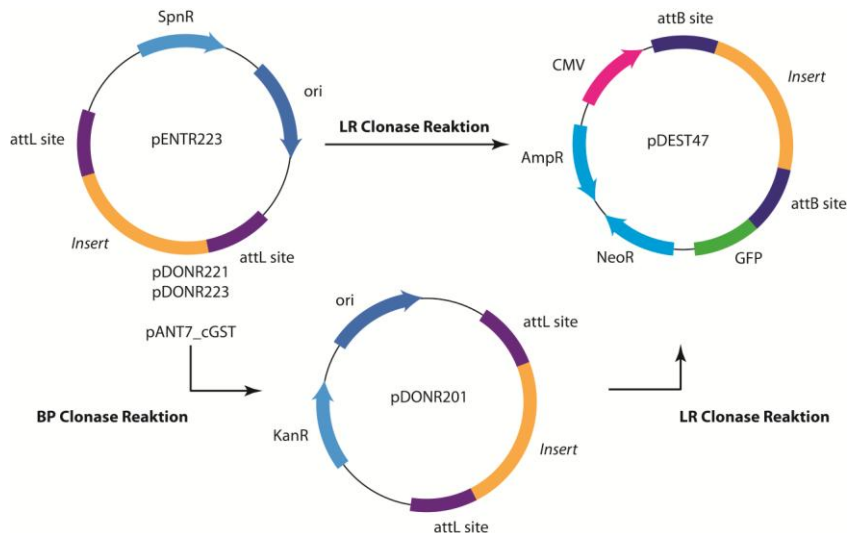


Abbildung 4.46: Schematische Darstellung der Vorgehensweise zur Erzeugung der gewünschten Expressionsklone für die Zielgene *HNRNPK*, *YBX1*, *SYNCRIP*, *GTF2B*, *HNRNPU*, *PCBP1* und *TOE1*.

Die Komponenten der BP-Reaktion (Tabelle) wurden zusammengegeben und die Reaktion 1 h bei 25 °C inkubiert. Um die Reaktion zu unterbrechen wurde Proteinase K (2 µL, 2 µg/µL) zugegeben und 10 min bei 37 °C inkubiert. Aus diesem Reaktionsgemisch wurde 1 µL entnommen und zu einem 50 µL Stock geeigneter Zellen (NEB 5-alpha, *New England BioLabs*) gegeben und chemisch transformiert. Der Rest der

Reaktionsmischung wurde bei -20 °C aufbewahrt um die Transformation wiederholen zu können.

Tabelle 4.5: Zusammensetzung der BP Clonase-Reaktion.

Komponente	Volumen [µL]
HNRNPU-pANT7_cGST (327 ng/µL)	1.5 µL
pDONR201 (150 ng/µL)	2.0 µL
5 × Clonase Reaction Buffer	4.0 µL
TE Puffer (pH 8.0)	8.5 µL
BP Clonase enzyme mix	4.0 µL

Die LR-Reaktionen wurden durchgeführt, indem ebenfalls alle Komponenten zusammengegeben wurden und die Reaktionen für 4 h bei 25 °C inkubiert wurden. Um die Reaktionen zu unterbrechen wurde Proteinase K (1 µL, 2 µg/µL) zugegeben und 10 min bei 37 °C inkubiert. Aus dieser Mischung wurde 1 µL entnommen und auf einen 50 µL Stock geeigneter Zellen (NEB 5-alpha, *New England BioLabs*) gegeben und chemisch transformiert. Der Rest der Reaktionsmischung wurde bei -20 °C aufbewahrt um die Transformation wiederholen zu können.

Tabelle 4.6: Zusammensetzung der LR Clonase-Reaktion

Komponente	Volumen [µL]
Entry Clone (100 ng/Reaktion)	0.27 µL – 0.84 µL
pDEST47 (150 ng/µL)	1.0 µL
TE Puffer (pH 8.0), <i>ad</i> 8.0 µL	6.16 µL – 6.73 µL
LR Clonase II enzyme mix	2.0 µL

DNA Restriktion

Um den Erfolg der Klonierungsreaktionen zu überprüfen wurden die erhaltenen Plasmide einem Restriktionsverdau unterzogen. Die DNA (400 ng/Probe) wurde in einem Gesamtvolumen von 10 µL in dem vom Hersteller empfohlenen Puffer für 1.5 h bis 3 h bei 37 °C verdaut und anschließend mittels Agarosegelelektrophorese analysiert.

Sequenzierung

Zusätzlich zu einem Restriktionsverdau wurden die erzeugten Plasmide zum Teil durch die Firma *GATC* (Konstanz) sequenziert. Hierzu wurde gemäß den Vorgaben des Anbieters 1.6 µg Plasmid (20 µL, 80 ng/µL) je Probe versandt. Die erhaltenen Sequenzen wurden mit *SerialCloner* (Version 2.6.1) auf Übereinstimmung mit der theoretischen Sequenz überprüft. Die verwendeten Primer sind in der unten stehenden Tabelle aufgeführt.

Tabelle 4.7: Für die Sequenzierung (*GATC*) von Plasmiden verwendete Primer.

Plasmid	Primer	Sequenz
HNRNPU-pANT7_cGST	<i>Custom</i>	5'-TGAAGGATGCCCAGAAG-3'
<i>Insert</i> -pENTR223	UP-40	5'-GTTTTCCCAGTCACGAC-3'
<i>Insert</i> -pDONR221	UP-40	5'-GTTTTCCCAGTCACGAC-3'
<i>Insert</i> -pDONR223	UP-40	5'-GTTTTCCCAGTCACGAC-3'
<i>Insert</i> -pDEST47	pc3.1GFP-Topo-RP	5'- CCCATTAACATCACC-3'
HNRNPK-pENTR223	M13-RP	5'-CAGGAAACAGCTATGACC-3'
HNRNPU-pANT7_cGST	pESP-RP	5'-TCCAAAAGAAGTCGAGTGG-3'
HNRNPU-pDONR201	pDONOR-RP	5'- GCAATGTAACATCAGAGAT-3'

Lokalisationsstudien

Für die subzelluläre Lokalisation der klonierten Binderproteine wurden GFP-Reporterplasmide verwendet. In einer 8-well Platte der Firma *Ibidi* (µslide) mit einer Oberfläche von 1 cm² wurden 3×10⁴ HeLa Zellen 24 h in 200 µL Standardmedium unter Normalbedingungen kultiviert. Mit Hilfe des Transfektionsreagenzes *JetPrime* der Firma *PeqLab* wurden die Zellen jeweils mit 1 µg Plasmid DNA in 3 µL Reaktionsmix transfiziert. Nach 4 h Inkubationszeit wurde das Medium ausgetauscht und die Zellen für weitere 20 h inkubiert. Nach 24 h sowie nach 48 h wurden die Zellen unter einem *Leica TCS SPE* Konfokalmikroskop betrachtet. Zur Visualisierung der Zellkerne wurde der Farbstoff *DRAQ5* (Thermo Fischer) verwendet. Als Anregungswellenlängen wurden für das GFP-Signal 488 nm und für den Farbstoff 358 nm verwendet.

Zellkultur Hap Zellen

Humane HAP1-Zellen (Horizon) und CDK5RAP1-Knockout Zellen (Horizon) wurden in Iscove's Modified Dulbecco's Medium (IMDM, Sigma-Aldrich) mit 10% FBS (Invitrogen), 4 mM L-Alanyl-L-Glutamin (200 mM, Sigma) und 1% Penicillin-Streptomycin (100%, Sigma) bei 37°C und 5% CO₂ kultiviert. Die Zellen wurden alle 2-3 Tage im Verhältnis 1:6 bzw. 1:12 passagiert. Die adhärenenten Zellen wurden in IMDM-Wachstumsmedium bis zu 80% Konfluenz bei 37°C und 5% CO₂ kultiviert, das Medium abgenommen und mit DPBS (Sigma) gewaschen. Anschließend wurden die Zellen mit Trypsin-EDTA Lösung (Gibco) versetzt und 5 min bei 37°C und 5% CO₂ inkubiert. Das Trypsin wurde mit IMDM-Medium inaktiviert, die Zellen vereinzelt und 3 min bei 300 xg pelletiert. Zur Langzeitlagerung in flüssigem Stickstoff wurden die Zellen in IMDM-Wachstumsmedium aufgenommen und die gleiche Menge (v/v) an Einfriermedium, bestehend aus 20% DMSO, 40% FBS und 40% IMDM, langsam zu gegeben. Die Zellen wurden 15 min bei RT äquilibriert und zur anschließenden Lagerung in *CryoTubes* aliquotiert.

Zelllysat Hap wt und CDK5RAP1 KO Zellen

Die humane HAP1-Zellen und CDK5RAP1-Knockout Zellen wurden mit 10 mL PBS gewaschen und anschließend 3 min bei 300 g pelletiert. Zu etwa 5×10^7 Zellen wurden 2.5 mL Extraktionspuffer (10mM Hepes KOH pH 7.9, 1.5mM MgCl₂, 10mM KCl) gegeben und 10 min auf Eis inkubiert. Anschließend wurde in einen Douncer überführt und 40 mal gedounced bis weniger als 50% lebende Zellen zu detektieren waren. Anschließend wurde bei 600 g 10 min bei 4 °C pelletiert. Im Pellet befinden sich die Zellkerne, die durch Zugabe von 200 µL Cell Lysis Puffer (10 mM Tris·HCl (pH 7.5), 150 mM NaCl, 0.5 mM EDTA, 0.5% (v/v) NP-40, 1 × Roche Complete Protease Inhibitor, 1% (v/v) Phosphatase Inhibitor Cocktail 2 und 3 (*Sigma Aldrich*)) für 30 min auf Eis lysiert wurden. Im Überstand befinden sich Cytosol und Mitochondrien, welche durch Pelettierung bei 11000 xg für 10 min bei 4 °C abgetrennt wurden. Im Pellet befinden sich nun die Mitochondrien welche durch Zugabe von 50 µL Lysis M Puffer mit Protease Inhibitor für 30 min auf Eis lysiert wurden.

Probenvorbereitung für die proteomweite Analyse

Für die proteomweite Analyse in Hap Wildtyp und CDK5RAP1 KO Zellen wurden 20 µg Zelllysate verwendet. Die Reduktion der Proben erfolgte durch Zugabe von 100 mM TCEP und einer Inkubation bei 60°C und 650rpm für 1 h. Anschließend wurden die Proben durch Zugabe von 200 mM Iodacetamid und Inkubation bei Raumtemperatur für 30 Minuten alkyliert. Der Verdau der Proben erfolgte mit 0.5 µg Trypsin (Promega) bei 37°C für 16 h. Um den Verdau zu beenden wurde anschließend 1 mM Phenylmethylsulfonylfluorid (PMSF) zugegeben. Die Aufreinigung der Proben erfolgte über StageTips.^[193]

LC-MS/MS Analyse

Die Proben wurden mit einem UltiMate 3000 RSLC Nano Flüssigkeitschromatographiesystem (Dionex, Thermo Fisher Scientific) und einem Q Exactive HF Massenspektrometer (Thermo Fisher Scientific) analysiert. Mit Hilfe einer µ-Vorfiltersäule (PepMap100, C₁₈, 5 µM, 100 Å, size 300 µm i.d. x 5 mm (Dionex, Thermo Fisher Scientific)) wurden die Proben erst aufkonzentriert und entsalzt und anschließend über eine selbstgemachte analytische Chromatographiesäule (ReproSil-Pur 120 C18-AQ, C₁₈, 1.9 µM, 120 Å (Dr. A. Maisch GmbH) in einem 75 µm i.d. x 150 mm Silica Picotip Emitter mit einem 8 µm Tip (New Objective) aufgetrennt. Die Proben wurden mit Hilfe eines 240 min dauernden Gradienten bei einer Flussrate von 300 nl/min aufgetrennt. Der Gradient wurde wie folgt verwendet:

1% für 3 Minuten; 1% - 8% über 2 Minuten; 8% - 32% über 195 Minuten; 32% - 60% über 15 Minuten; 60% - 85% über 5 Minuten; 85% für 10 Minuten; 85% - 1% über 5 Minuten; 1% für 5 Minuten.

Es wurden ausschließlich LC-MS *grade* Lösungsmittel verwendet. Puffer A enthält Wasser und 0.1% Ameisensäure, Puffer B enthält Acetonitril und 0.1% Ameisensäure.

Die massenspektrometrische Analyse erfolgte im *Full Mass Scan* zwischen den Grenzen m/z 300 und m/z 1750 bei einer Auflösung von 120000 und einer maximalen Ioneninjektionszeit von 120 ms im *profile* Modus. Diesem Suchscan folgend wurden die 20 intensivsten Ionen selektiert, fragmentiert und im *centroid* Modus mit den folgenden

Parametern gemessen: einer Auflösung von 30000, ein AGC Target von 2×10^5 Ladungen, einer maximalen Ioneninjektionszeit von 70 ms, einem Isolationsfenster von 2 m/z mit einem Abstand von +0.4 m/z und einer normalisierten HCD Energie von 28%.

Signale mit unbekannten Ladungszustand oder einem Ladungszustand von 1, 7, 8 oder höher wurden nicht für die Fragmentierung ausgewählt. Um *supersampling* zu vermeiden, wurden die Signale für 40 s ausgeschlossen, nachdem sie für eine Isolation und Fragmentierung selektiert wurden. Die *Peptide-match* Einstellung wurde auf "bevorzugt" gestellt und die *exclude isotope* Einstellung wurde auf "an" gestellt.

LFQ Daten Auswertung

Für die Auswertung durch Label freie Quantifizierung wurde die Software MaxQuant (Version 1.5.0.25) verwendet. Die Quantifizierung erfolgte unter Verwendung von drei biologischen Replikaten.

Die Andromeda Suchmaschine wurde in Kombination mit Uniprot Datenbanken verwendet. Es wurden maximal zwei *missed cleavage sites* erlaubt. Die *peptide tolerance* während der Hauptsuche wurde auf 4.5 ppm gestellt. Carbamidomethyl (C) wurde als statische Modifikation eingestellt. Variable Modifikationen waren Acetylgruppen (Protein N-term) und Oxidationen (M). Der LFQ Algorithmus wurde mit *default* Einstellungen angewendet. Die Möglichkeit "*match between runs*" wurde auch verwendet.

Die LFQ Daten wurden anschließend mit der Perseus Software (Version 1.5.0.9) analysiert. Die LFQ Intensitäten wurden logarithmiert und es wurden nur Proteine identifiziert die in Allen 3 Replikaten zu finden waren.

GO Analyse

Die *Gene Ontology* Analyse der Daten erfolgte mit Datenbanken für Annotierung Visualisierung und Integrated Discovery (DAVID Bioinformatics Resources 6.7).

5 Abkürzungsverzeichnis

-/-	Knockout beider Allele eines Enzyms
+/-	Knockout eines der beiden Allele eines Enzyms
5caC	5-Carboxycytosin
5cadC	5-Carboxy-2'-desoxycytidin
5cadU	5-Carboxy-2'-desoxyuridin
5caU	5-Carboxyuracil
5fC	5-Formylcytosin
5fdC	5-Formyl-2'-desoxycytidin
5fdU	5-Formyl-2'-desoxyuridin
5fU	5-Formyluracil
5hmC	5-Hydroxymethylcytosin
5hmdC	5-Hydroxymethyl-2'-desoxycytidin
5hmdU	5-Hydroxymethyl-2'-desoxyuridin
5hmU	5-Hydroxymethyluracil
5mC	5-Methylcytosin
5mdC	5-Methyl-2'-desoxycytidin
8oxodG	8-Oxo-2'-desoxyguanosin
8oxoG	8-Oxoguanin
A	Adenin
Ac	Acetyl
AID	<i>activation-induced deaminase</i>
amu	<i>atomic mass unit</i>
Apobec	<i>apolipoprotein B pre-mRNA editing enzyme, catalytic polypeptide</i>
AP-Stelle	apurinische/apyrimidinische Stelle
APC	<i>Adenomatous-polyposis-coli</i> Protein
ATP	Adenosintriphosphat
BER	Basenexzisionsreparatur
C	Cytosin
Cda	Cytosin-Desaminase
CDK5RAP1	<i>CDK5 regulatory subunit associated protein 1</i>
CE	<i>collision energy</i>

CID	<i>collision induced dissociation</i>
CpG	dC-dG-Dinukleotid
dA	2'-Desoxyadenosin
DAPI	4',6-Diamidin-2-phenylindol
dC	2'-Desoxycytidin
DC	Dünnschichtchromatographie
DCM	Dichlormethan
Dctd	Desoxycytidilat-Desaminase
dG	2'-Desoxyguanosin
dN	2'-Desoxynukleosid
DNA	<i>deoxyribonucleic acid</i> ; Desoxyribonukleinsäure
Dnmt	DNA-Methyltransferase
dsDNA	doppelsträngige DNA
dT	2'-Desoxythymidin
dTMP	2'-Desoxythymidin-Monophosphat
dTTP	2'-Desoxythymidin-Triphosphat
dU	2'-Desoxyuridin
EDTA	<i>ethylenediaminetetraacetic acid</i>
EpiLC	<i>epiblast like stem cells</i>
ESI	Elektrospray-Ionisierung
ES-Zelle	embryonale Stammzelle
Et	Ethyl
EtOAc	Ethylacetat
EtOH	Ethanol
FACS	<i>fluorescence activated cell sorting</i>
FAD	Flavin-Adenin-Dinukleotid
FT-ICR	<i>Fourier-Transformation-Ion-Cyclotron</i>
G	Guanin
GFP	Grün fluoreszierendes Protein
GTF2B	Genereller Transkriptionsfaktor 2B
H3K14ac	Acetyliertes Lysin 14 in Histon H3
H3K27me3	Trimethyliertes Lysin 27 in Histon H3
H3K9ac	Acetyliertes Lysin 9 in Histon H3

H3K9me3	Trimethyliertes Lysin 9 in Histon H3
HAT	Histon-Acetyltransferase
HDAC	Histon-Deacetylase
HEK	<i>human embryoid kidney</i>
HeLa	humane Zervix Adenokarzinom Zelllinie
Hepes	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
HMEC	humane mikrovaskuläre Endothelzelllinie
HMT	Histon-Methyltransferase
HNRNPK	<i>heterogeneous nuclear ribonucleoprotein K</i>
HNRNPU	<i>heterogeneous nuclear ribonucleoprotein U</i>
HP1	Heterochromatin-Protein 1
HPLC	Hochdruckflüssigkeitschromatographie
HRMS	hochauflösende (<i>high resolution</i>) Massenspektrometrie
i ⁶ A	N ⁶ -Isopentenyladenosin
Idc	Isoorotatdecarboxylase
Idh	Isocitrat-Dehydrogenase
iPrOH	Isopropanol
iPS-Zellen	induzierte pluripotente Stammzellen
Isoorotat	5-Carboxyuracil (5caU)
IZM	innere Zellmasse
KD	<i>Knockdown</i>
KO	<i>Knockout</i>
LIF	<i>leukemia inhibitory factor</i>
m ⁶ A	N ⁶ -Methyladenosin
m ⁷ G	N ⁷ -Methylguanosin
ms ² i ⁶ A	2-Methylthio-N ⁶ -isopentenyladenosin
m ⁵ C	5-Methylcytosin
MALDI-TOF	<i>matrix assisted laser desorption ionization time of flight</i>
MCF-7	humane Brustkrebs Zelllinie
Me	Methyl
MeCN	Acetonitril
MeCP2	<i>methyl-CpG binding protein 2</i>
MeOH	<i>Methanol</i>

mES-Zelle	embryonale Stammzelle der Maus
MLL	<i>Mixed Lineage Leukemia</i>
MS	Massenspektrometrie, Massenspektrometer
MS/MS	Tandemmassenspektrometrie
Nanog	Transkriptionsfaktor
Neil	<i>endonuclease VIII-like glycosylase</i>
NMR	<i>nucleic magnetic resonance</i>
Oct4	Octamer-bindender Transkriptionsfaktor 4
ODN	Oligodesoxynukleotid
OG	2-Oxoglutarat
Ogt	<i>O-linked β-D-N-Acetylglucosamin-Transferase</i>
P	Signifikanzwert berechnet durch einen <i>t</i> -Test
PAGE	Polyacrylamidgelelektrophorese
PBS	<i>phosphate buffered saline</i>
PDB	<i>protein data bank</i>
Q	Queuosin
ROS	<i>reactive oxygen species</i>
Rpm	<i>rounds per minute</i>
SAM	S-Adenosylmethionin
SD	Standardabweichung; <i>standard deviation</i>
SILAC	<i>stable isotope labeling by amino acids in cell culture</i>
SINCRIP	<i>synaptotagmin binding cytoplasmic RNA interacting protein</i>
Smug1	<i>single-stranded-selective monofunctional uracil-DNA glycosylase 1</i>
ssDNA	einzelsträngige (<i>single-stranded</i>) DNA
T	Thymin
t ⁶ A	N ⁶ -Threonylcarbamoyladenosen
TDG	Thymin-DNA-Glykosylase
Tet	<i>Ten-eleven translocation</i> Protein
TFAM	<i>transcription factor A, mitochondrial</i>
U	Uracil
U2AF2	<i>U2 small nuclear RNA auxiliary factor 2</i>
UDG	Uracil-DNA-Glykosylase
UHPLC	Ultrahochdruckflüssigkeitschromatographie

UHPLC -ESIMS/MS	Ultrahochdruckflüssigkeitschromatographie, Elektrospray-Ionisierung und Tandemmassenspektrometrie
Uhrf	<i>ubiquitin-like, containing PHD and RING finger domains</i>
UV	ultraviolettes Licht
v/v	Volumenanteil
YBX1	<i>Y-box binding protein 1</i>
yW	Wybutosin

6 Literaturverzeichnis

- [1] A. D. Hershey, M. Chase, *J. Gen. Physiol.* **1952**, *36*, 39-56. Independent functions of viral protein and nucleic acid in growth of bacteriophage
- [2] O. T. Avery, C. M. Macleod, M. McCarty, *J. Exp. Med.* **1944**, *79*, 137-158. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III
- [3] R. D. Hotchkiss, *J. Biol. Chem.* **1948**, *175*, 315-332. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography
- [4] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1
- [5] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem., Int. Ed.* **2011**, *50*, 7008-7012. The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA
- [6] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine
- [7] M. Ehrlich, R. Wang, *Science* **1981**, *212*, 1350-1357. 5-Methylcytosine in eukaryotic DNA
- [8] M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **2005**, *74*, 481-514. Eukaryotic Cytosine Methyltransferases
- [9] R. Metivier, R. Gallais, C. Tiffocche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, *Nature* **2008**, *452*, 45-50. Cyclical DNA methylation of a transcriptionally active promoter
- [10] W. Reik, *Nature* **2007**, *447*, 425-432. Stability and flexibility of epigenetic gene regulation in mammalian development
- [11] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41-45. The language of covalent histone modifications
- [12] K. Tony, *Cell* **2007**, *128*, 693-705. Chromatin Modifications and Their Function
- [13] V. A. Spencer, J. R. Davie, *Gene* **1999**, *240*, 1-12. Role of covalent modifications of histones in regulating gene expression
- [14] A. J. Bannister, R. Schneider, T. Kouzarides, *Cell* **2002**, *109*, 801-806. Histone Methylation: Dynamic or Static?
- [15] B. M. Turner, L. P. O'Neill, *Seminars in Cell Biology* **1995**, *6*, 229-236. Histone acetylation in chromatin and chromosomes
- [16] P. Loidl, *Chromosoma* **1994**, *103*, 441-449. Histone acetylation: facts and questions
- [17] Z. Liutkevičiūtė, G. Lukinavičius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400-402. Cytosine-5-methyltransferases add aldehydes to DNA
- [18] M. Okano, S. Xie, E. Li, *Nat. Genet.* **1998**, *19*, 219-220. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases
- [19] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development

- [20] T. H. Bestor, V. M. Ingram, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 5559-5563. Two DNA methyltransferases from murine erythroleukemia cells: Purification, sequence specificity, and mode of interaction with DNA
- [21] T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, *J. Mol. Biol.* **1988**, *203*, 971-983. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases
- [22] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204-220. Establishing, maintaining and modifying DNA methylation patterns in plants and animals
- [23] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315-322. Human DNA methylomes at base resolution show widespread epigenomic differences
- [24] J. U. Guo, Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G. L. Ming, H. Song, *Nat. Neurosci.* **2014**, *17*, 215-222. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain
- [25] R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghghi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, *Science* **2013**, *341*, 1237905. Global epigenomic reconfiguration during mammalian brain development
- [26] F. Larsen, G. Gundersen, R. Lopez, H. Prydz, *Genomics* **1992**, *13*, 1095-1107. CpG islands as gene markers in the human genome
- [27] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204-220. Establishing, maintaining and modifying DNA methylation patterns in plants and animals
- [28] M. Ehrlich, M. A. Gama-Sosa, L.-H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nucleic Acids Res.* **1982**, *10*, 2709-2721. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells
- [29] M. W. Simmen, *Genomics* **2008**, *92*, 33-40. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals
- [30] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBiochem* **2011**, *12*, 206-222. Structure and function of mammalian DNA methyltransferases
- [31] B. K. Duncan, J. H. Miller, *Nature* **1980**, *287*, 560-561. Mutagenic deamination of cytosine residues in DNA
- [32] F. Santos, J. Peat, H. Burgess, C. Rada, W. Reik, W. Dean, *Epigenetics Chromatin* **2013**, *6*, 39. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair
- [33] M. Gehring, W. Reik, S. Henikoff, *Trends Genet.* **2009**, *25*, 82-90. DNA demethylation by DNA repair
- [34] T. Isobe, S. N. Song, P. Tiwari, H. Ito, Y. Yamaguchi, K. Yoshizaki, *FEBS Lett.* **2013**, *587*, 2487-2492. Activation-induced cytidine deaminase auto-activates and triggers aberrant gene expression
- [35] R. Kumar, L. DiMenna, N. Schrode, T. C. Liu, P. Franck, S. Munoz-Descalzo, A. K. Hadjantonakis, A. A. Zarrin, J. Chaudhuri, O. Elemento, T. Evans, *Nature* **2013**, *500*, 89-92. AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes
- [36] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W.

- Reik, *Nature* **2010**, *463*, 1101-1105. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency
- [37] M. Weber, I. Hellmann, M. B. Stadler, L. Ramos, S. Pääbo, M. Rebhan, D. Schübeler, *Nat. Genet.* **2007**, *39*, 457-466. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome
- [38] A. P. Feinberg, *Nature* **2007**, *447*, 433-440. Phenotypic plasticity and the epigenetics of human disease
- [39] P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068-1070. The role of DNA methylation in mammalian epigenetics
- [40] A. P. Feinberg, B. Vogelstein, *Nature* **1983**, *301*, 89-92. Hypomethylation distinguishes genes of some human cancers from their normal counterparts
- [41] F. Gaudet, J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, R. Jaenisch, *Science* **2003**, *300*, 489-492. Induction of Tumors in Mice by Genomic Hypomethylation
- [42] A. T. De Souza, T. Yamada, J. J. Mills, R. L. Jirtle, *FASEB J.* **1997**, *11*, 60-67. Imprinted genes in liver carcinogenesis
- [43] E. Giovannucci, *Horm. Metab. Res.* **2003**, *35*, 694-704. Nutrition, insulin, insulin-like growth factors and cancer
- [44] A. Kaneda, A. P. Feinberg, *Cancer Res.* **2005**, *65*, 11236-11240. Loss of Imprinting of IGF2: A Common Epigenetic Modifier of Intestinal Tumor Risk
- [45] E. N. Gal-Yam, Y. Saito, G. Egger, P. A. Jones, *Annu. Rev. Med.* **2008**, *59*, 267-280. Cancer epigenetics: modifications, screening, and therapy
- [46] T. K. Kelly, C. D. D. De, P. A. Jones, *Nat. Biotechnol.* **2010**, *28*, 1069-1078. Epigenetic modifications as therapeutic targets
- [47] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930. The nuclear DNA base 5-Hydroxymethylcytosine is present in Purkinje neurons and the brain
- [48] G. R. Wyatt, S. S. Cohen, *Biochem. J.* **1953**, *55*, 774-782. Bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine
- [49] M. Muenzel, D. Globisch, T. Carell, *Angew. Chem., Int. Ed.* **2011**, *50*, 6460-6468. 5-Hydroxymethylcytosine, the Sixth Base of the Genome
- [50] C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580-583. Oxygenase catalyzed 5-methylcytosine hydroxylation
- [51] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129-1133. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification
- [52] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781-790. Presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid
- [53] D. Globisch, M. Muenzel, M. Mueller, S. Michalakakis, M. Wagner, S. Koch, T. Brueckl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates
- [54] M. Muenzel, D. Globisch, T. Brueckl, M. Wagner, V. Welzmueller, S. Michalakakis, M. Mueller, M. Biel, T. Carell, *Angew. Chem., Int. Ed.* **2010**, *49*, 5375-5377, S5375/5371-S5375/5324. Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain
- [55] M. Muenzel, **2011**. Synthesis of Natural DNA Modifications and Their Detection in DNA-Identification and Total Synthesis of a Novel DNA Photolesion
- [56] T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Munzel, M. Muller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schuller, T. Carell, H. A.

- Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577-1590. Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors
- [57] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, *Science* **2011**, *333*, 1303-1307. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA
- [58] C. X. Song, C. He, *Trends Biochem. Sci.* **2013**, *38*, 480-484. Potential functional roles of DNA demethylation intermediates
- [59] O. Abdel-Wahab, A. Mullally, C. Hedvat, G. Garcia-Manero, J. Patel, M. Wadleigh, S. Malinger, J. Yao, O. Kilpivaara, R. Bhat, K. Huberman, S. Thomas, I. Dolgalev, A. Heguy, E. Paietta, M. M. Le Beau, M. Beran, M. S. Tallman, B. L. Ebert, H. M. Kantarjian, R. M. Stone, D. G. Gilliland, J. D. Crispino, R. L. Levine, *Blood* **2009**, *114*, 144-147. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies
- [60] F. Delhommeau, S. Dupont, V. Della Valle, C. James, S. Trannoy, A. Masse, O. Kosmider, J. P. Le Couedic, F. Robert, A. Alberdi, Y. Lecluse, I. Plo, F. J. Dreyfus, C. Marzac, N. Casadevall, C. Lacombe, S. P. Romana, P. Dessen, J. Soulier, F. Viguie, M. Fontenay, W. Vainchenker, O. A. Bernard, *N. Engl. J. Med.* **2009**, *360*, 2289-2301. Mutation in TET2 in myeloid cancers
- [61] S. M. Langemeijer, R. P. Kuiper, M. Berends, R. Knops, M. G. Aslanyan, M. Massop, E. Stevens-Linders, P. van Hoogen, A. G. van Kessel, R. A. Raymakers, E. J. Kamping, G. E. Verhoef, E. Verburch, A. Hagemeijer, P. Vandenberghe, T. de Witte, B. A. van der Reijden, J. H. Jansen, *Nat. Genet.* **2009**, *41*, 838-842. Acquired mutations in TET2 are common in myelodysplastic syndromes
- [62] C. G. Mullighan, *Nat. Genet.* **2009**, *41*, 766-767. TET2 mutations in myelodysplasia and myeloid malignancies
- [63] M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839-843. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2
- [64] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692-706. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics
- [65] M. Xiao, H. Yang, W. Xu, S. Ma, H. Lin, H. Zhu, L. Liu, Y. Liu, C. Yang, Y. Xu, S. Zhao, D. Ye, Y. Xiong, K. L. Guan, *Genes Dev.* **2012**, *26*, 1326-1338. Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors
- [66] M. E. Figueroa, O. Abdel-Wahab, C. Lu, P. S. Ward, J. Patel, A. Shih, Y. Li, N. Bhagwat, A. Vasanthakumar, H. F. Fernandez, M. S. Tallman, Z. Sun, K. Wolniak, J. K. Peeters, W. Liu, S. E. Choe, V. R. Fantin, E. Paietta, B. Lowenberg, J. D. Licht, L. A. Godley, R. Delwel, P. J. Valk, C. B. Thompson, R. L. Levine, A. Melnick, *Cancer Cell* **2010**, *18*, 553-567. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation
- [67] W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S.-H. Kim, S. Ito, C. Yang, P. Wang, M.-T. Xiao, *Cancer Cell* **2011**, *19*, 17-30. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases
- [68] L. Cimmino, M. M. Dawlaty, D. Ndiaye-Lobry, Y. S. Yap, S. Bakogianni, Y. Yu, S.

- Bhattacharyya, R. Shaknovich, H. Geng, C. Lobry, J. Mullenders, B. King, T. Trimarchi, B. Aranda-Orgilles, C. Liu, S. Shen, A. K. Verma, R. Jaenisch, I. Aifantis, *Nat. Immunol.* **2015**, *16*, 653-662. TET1 is a tumor suppressor of hematopoietic malignancy
- [69] H. Yang, Y. Liu, F. Bai, J. Y. Zhang, S. H. Ma, J. Liu, Z. D. Xu, H. G. Zhu, Z. Q. Ling, D. Ye, K. L. Guan, Y. Xiong, *Oncogene* **2013**, *32*, 663-669. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation
- [70] E. Meshorer, T. Misteli, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 540-546. Chromatin in pluripotent embryonic stem cells and differentiation
- [71] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607-620. Active DNA demethylation: many roads lead to Rome
- [72] H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45-68. Reversing DNA methylation: mechanisms, genomics, and biological functions
- [73] J. Nichols, A. Smith, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a008128. Pluripotency in the embryo and in culture
- [74] A. Meissner, *Nat. Biotechnol.* **2010**, *28*, 1079-1088. Epigenetic modifications in pluripotent and differentiated cells
- [75] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502. Demethylation of the zygotic paternal genome
- [76] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Curr. Biol.* **2000**, *10*, 475-478. Active demethylation of the paternal genome in the mouse zygote
- [77] F. Santos, B. Hendrich, W. Reik, W. Dean, *Dev. Biol.* **2002**, *241*, 172-182. Dynamic reprogramming of DNA methylation in the early mouse embryo
- [78] M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *Nat. Commun.* **2011**, *2*, 241. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming
- [79] H. Fulka, M. Mrazek, O. Tepla, J. Fulka, Jr., *Reproduction* **2004**, *128*, 703-708. DNA methylation pattern in human zygotes and developing embryos
- [80] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T.-P. Gu, B. Hu, Colum P. Walsh, J. Li, F. Tang, G.-L. Xu, *Cell Stem Cell* **2014**, *15*, 447-458. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote
- [81] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T.-P. Gu, B. Hu, Colum P. Walsh, J. Li, F. Tang, G.-L. Xu, *Cell Stem Cell*, *15*, 447-458. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote
- [82] L. Shen, A. Inoue, J. He, Y. Liu, F. Lu, Y. Zhang, *Cell Stem Cell* **2014**, *15*, 459-470. Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes
- [83] A. Inoue, L. Shen, S. Matoba, Y. Zhang, *Cell Rep.* **2015**, *10*, 463-470. Haploinsufficiency, but Not Defective Paternal 5mC Oxidation, Accounts for the Developmental Defects of Maternal Tet3 Knockouts
- [84] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Muller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem., Int. Ed.* **2012**, *51*, 6516-6520. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing
- [85] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schroder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599. Deamination, oxidation, and C-C

- bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine
- [86] S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G. L. Xu, J. Ding, *Cell Res.* **2013**, *23*, 1296-1309. Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase
- [87] Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute, S. Klimasauskas, *J. Am. Chem. Soc.* **2014**, *136*, 5884-5887. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases
- [88] A. Liu, H. Zhang, *Biochemistry* **2006**, *45*, 10407-10411. Transition metal-catalyzed nonoxidative decarboxylation reactions
- [89] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, A. A. Axhemi, *Biochim. Biophys. Acta* **2005**, *1723*, 256-264. Genes of the thymidine salvage pathway: Thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*
- [90] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential implications for active demethylation of CpG Sites
- [91] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Muller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Muller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives
- [92] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, C. M. Le, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67-79. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair
- [93] H. Hashimoto, S. Hong, A. S. Bhagwat, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, *40*, 10203-10214. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation
- [94] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, advance online publication. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA
- [95] U. Muller, C. Bauer, M. Siegl, A. Rottach, H. Leonhardt, *Nucleic Acids Res.* **2014**, *42*, 8592-8604. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation
- [96] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *J. Am. Chem. Soc.* **2013**, *135*, 15813-15822. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA
- [97] H. Hashimoto, X. Zhang, X. Cheng, *DNA Repair (Amst)* **2013**, *12*, 535-540. Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH
- [98] A. Maiti, A. C. Drohat, *DNA Repair (Amst)* **2011**, *10*, 545-553. Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA

- glycosylase: Insights into recognition and processing of G.T mispairs
- [99] L. Schomacher, D. Han, M. U. Musheev, K. Arab, S. Kienhofer, A. von Seggern, C. Niehrs, *Nat. Struct. Mol. Biol.* **2016**, *advance online publication*. Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation
- [100] M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schoeler, J. Walter, *EMBO J.* **2010**, *29*, 1877-1888. Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes
- [101] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**, *279*, 52353-52360. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming
- [102] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *Cell* **2008**, *135*, 1201-1212. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45
- [103] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Nature* **2010**, *463*, 1042-1047. Reprogramming towards pluripotency requires AID-dependent DNA demethylation
- [104] F. L. Chung, H. J. Chen, R. G. Nath, *Carcinogenesis* **1996**, *17*, 2105-2111. Lipid peroxidation as a potential endogenous source for the formation of exocyclic DNA adducts
- [105] F. el Ghissassi, A. Barbin, J. Nair, H. Bartsch, *Chem. Res. Toxicol.* **1995**, *8*, 278-283. Formation of 1,N6-ethenoadenine and 3,N4-ethenocytosine by lipid peroxidation products and nucleic acid bases
- [106] L. Gros, A. A. Ishchenko, M. Saparbaev, *Mutat. Res.* **2003**, *531*, 219-229. Enzymology of repair of etheno-adducts
- [107] J. Nair, A. Barbin, Y. Guichard, H. Bartsch, *Carcinogenesis* **1995**, *16*, 613-617. 1,N6-ethenodeoxyadenosine and 3,N4-ethenodeoxycytine in liver DNA from humans and untreated rodents detected by immunoaffinity/32P-postlabeling
- [108] D. Zdzalik, A. Domanska, P. Prorok, K. Kosicki, E. van den Born, P. O. Falnes, C. J. Rizzo, F. P. Guengerich, B. Tudek, *DNA Repair (Amst)* **2015**, *30*, 1-10. Differential repair of etheno-DNA adducts by bacterial and human AlkB proteins
- [109] B. I. Fedeles, V. Singh, J. C. Delaney, D. Li, J. M. Essigmann, *J. Biol. Chem.* **2015**, *290*, 20734-20742. The AlkB Family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond
- [110] J. C. Delaney, L. Smeester, C. Wong, L. E. Frick, K. Taghizadeh, J. S. Wishnok, C. L. Drennan, L. D. Samson, J. M. Essigmann, *Nat. Struct. Mol. Biol.* **2005**, *12*, 855-860. AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo
- [111] A. M. Maciejewska, K. P. Ruszel, J. Nieminuszczy, J. Lewicka, B. Sokolowska, E. Grzesiuk, J. T. Kusmierk, *Mutat. Res.* **2010**, *684*, 24-34. Chloroacetaldehyde-induced mutagenesis in *Escherichia coli*: the role of AlkB protein in repair of 3,N(4)-ethenocytosine and 3,N(4)-alpha-hydroxyethanocytosine
- [112] K. Linhart, H. Bartsch, H. K. Seitz, *Redox. Biol.* **2014**, *3*, 56-62. The role of reactive oxygen species (ROS) and cytochrome P-450 2E1 in the generation of carcinogenic etheno-DNA adducts
- [113] L. M. Coussens, Z. Werb, *Nature* **2002**, *420*, 860-867. Inflammation and cancer
- [114] H. Ohshima, H. Bartsch, *Mutat. Res.* **1994**, *305*, 253-264. Chronic infections and inflammatory processes as cancer risk factors: possible role of nitric oxide in

- carcinogenesis
- [115] S. Suerbaum, P. Michetti, *N. Engl. J. Med.* **2002**, *347*, 1175-1186. *Helicobacter pylori* infection
 - [116] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nat. Protoc.* **2008**, *3*, 1287-1298. Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry
 - [117] P. C. Dedon, S. R. Tannenbaum, *Arch. Biochem. Biophys.* **2004**, *423*, 12-22. Reactive nitrogen species in the chemical biology of inflammation
 - [118] M. Shen, P. Bin, H. Li, X. Zhang, X. Sun, H. Duan, Y. Niu, T. Meng, Y. Dai, W. Gao, S. Yu, G. Gu, Y. Zheng, *Sci. Total Environ.* **2016**, *543*, 267-273. Increased levels of etheno-DNA adducts and genotoxicity biomarkers of long-term exposure to pure diesel engine exhaust
 - [119] C. C. Garcia, G. L. Batista, F. P. Freitas, F. S. Lopes, A. B. Sanchez, I. G. Gutz, P. Di Mascio, M. H. Medeiros, *Free Radic. Biol. Med.* **2014**, *75 Suppl 1*, S41. Quantification of DNA adducts in lungs, liver and brain of rats exposed to acetaldehyde
 - [120] W. E. Cohn, E. Volkin, *Nature* **1951**, *167*, 483-484. Nucleoside-5[prime]-Phosphates from Ribonucleic Acid
 - [121] C. T. Yu, F. W. Allen, *Biochim. Biophys. Acta* **1959**, *32*, 393-406. Studies on an isomer of uridine isolated from ribonucleic acids
 - [122] M. Charette, M. W. Gray, *IUBMB Life* **2000**, *49*, 341-351. Pseudouridine in RNA: what, where, how, and why
 - [123] P. A. Limbach, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1994**, *22*, 2183-2196. Summary: the modified nucleosides of RNA
 - [124] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195-D201. The RNA modification database, RNAMDB: 2011 update
 - [125] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262-267. MODOMICS: a database of RNA modification pathways--2013 update
 - [126] E. M. Phizicky, J. D. Alfonzo, *FEBS Lett.* **2010**, *584*, 265-271. Do all modifications benefit all tRNAs?
 - [127] V. Reiter, **2013**. Quantifizierung eukaryotischer RNA-Modifikationen und Untersuchungen zum humanen Homolog der bakteriellen Methylthiotransferase MiaB
 - [128] P. F. Agris, F. A. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1-13. tRNA's wobble decoding of the genome: 40 years of modification
 - [129] H. Grosjean, M. Sprinzl, S. Steinberg, *Biochimie* **1995**, *77*, 139-141. Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies
 - [130] D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Hershkovitz, D. Han, L. C. Doré, N. Amariglio, G. Rechavi, C. He, *Nature* **2016**, *530*, 441-446. The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA
 - [131] I. G. Laptev, A. Y. Golovina, P. V. Sergiev, O. A. Dontsova, *Mol. Biol.* **2015**, *49*, 923-

936. Posttranscriptional messenger RNA modifications in eukaryotes
- [132] J. Karjane, C. Yi, Y. T. Yu, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 581-585. Transcriptome-wide dynamics of RNA pseudouridylation
- [133] I. A. Roundtree, C. He, *Curr. Opin. Chem. Biol.* **2016**, *30*, 46-51. RNA epigenetics-chemical messages for posttranscriptional gene regulation
- [134] K. D. Meyer, S. R. Jaffrey, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 313-326. The dynamic epitranscriptome: N6-methyladenosine and gene expression control
- [135] A. Maity, B. Das, *FEBS J.* **2015**. N6-methyladenosine modification in mRNA: machinery, function and implications for health and diseases
- [136] K. H. Kaminska, U. Baraniak, M. Boniecki, K. Nowaczyk, A. Czerwonec, J. M. Bujnicki, *Proteins* **2008**, *70*, 1-18. Structural bioinformatics analysis of enzymes involved in the biosynthesis pathway of the hypermodified nucleoside ms(2)io(6)A37 in tRNA
- [137] J. A. Moore, C. D. Poulter, *Biochemistry* **1997**, *36*, 604-614. Escherichia coli dimethylallyl diphosphate:tRNA dimethylallyltransferase: a binding mechanism for recombinant enzyme
- [138] Y. Motorin, G. Bec, R. Tewari, H. Grosjean, *Nat Biotechnol* **1997**, *3*, 721-733. Transfer RNA recognition by the Escherichia coli delta2-isopentenyl-pyrophosphate:tRNA delta2-isopentenyl transferase: dependence on the anticodon arm structure
- [139] T. Soderberg, C. D. Poulter, *Biochemistry* **2000**, *39*, 6546-6553. Escherichia coli dimethylallyl diphosphate:tRNA dimethylallyltransferase: essential elements for recognition of tRNA substrates within the anticodon stem-loop
- [140] S. Chimnaronk, F. Forouhar, J. Sakai, M. Yao, C. M. Tron, M. Atta, M. Fontecave, J. F. Hunt, I. Tanaka, *Biochemistry* **2009**, *48*, 5057-5065. Snapshots of dynamics in synthesizing N(6)-isopentenyladenosine at the tRNA anticodon
- [141] A. Golovko, G. Hjälm, F. Sitbon, B. Nicander, *Gene* **2000**, *258*, 85-93. Cloning of a human tRNA isopentenyl transferase
- [142] F. Pierrel, T. Douki, M. Fontecave, M. Atta, *J. Biol. Chem.* **2004**, *279*, 47555-47563. MiaB protein is a bifunctional radical-S-adenosylmethionine enzyme involved in thiolation and methylation of tRNA
- [143] A. H. Rosenberg, M. L. Gefter, *J. Mol. Biol.* **1969**, *46*, 581-584. An iron-dependent modification of several transfer RNA species in Escherichia coli
- [144] M. Buck, E. Griffiths, *Nucleic Acids Res.* **1982**, *10*, 2609-2624. Iron mediated methylthiolation of tRNA as a regulator of operon expression in Escherichia coli
- [145] M. Fontecave, S. Ollagnier-de-Choudens, E. Mulliez, *Chem. Rev.* **2003**, *103*, 2149-2166. Biological radical sulfur insertion reactions
- [146] H. L. Hernandez, F. Pierrel, E. Elleingand, R. Garcia-Serres, B. H. Huynh, M. K. Johnson, M. Fontecave, M. Atta, *Biochemistry* **2007**, *46*, 5140-5147. MiaB, a bifunctional radical-S-adenosylmethionine enzyme involved in the thiolation and methylation of tRNA, contains two essential [4Fe-4S] clusters
- [147] V. Reiter, D. M. S. Matschkal, M. Wagner, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, *40*, 6235-6240. The CDK5 repressor CDK5RAP1 is a methylthiotransferase acting on nuclear and mitochondrial RNA
- [148] S. Arragain, S. K. Handelman, F. Forouhar, F. Y. Wei, K. Tomizawa, J. F. Hunt, T. Douki, M. Fontecave, E. Mulliez, M. Atta, *J. Biol. Chem.* **2010**, *285*, 28425-28433. Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N6-threonylcarbamoyladenine in tRNA
- [149] F. Y. Wei, B. Zhou, T. Suzuki, K. Miyata, Y. Ujihara, H. Horiguchi, N. Takahashi, P. Xie,

- H. Michiue, A. Fujimura, T. Kaitsuka, H. Matsui, Y. Koga, S. Mohri, T. Suzuki, Y. Oike, K. Tomizawa, *Cell Metab.* **2015**, *21*, 428-442. Cdk5rap1-mediated 2-methylthio modification of mitochondrial tRNAs governs protein translation and contributes to myopathy in mice and humans
- [150] I. Diaz, M. Ehrenberg, *J. Mol. Biol.* **1991**, *222*, 1161-1171. ms2i6A deficiency enhances proofreading in translation
- [151] M. L. Gefter, R. L. Russell, *J. Mol. Biol.* **1969**, *39*, 145-157. Role modifications in tyrosine transfer RNA: a modified base affecting ribosome binding
- [152] J. Cabello-Villegas, M. E. Winkler, E. P. Nikonowicz, *J. Mol. Biol.* **2002**, *319*, 1015-1034. Solution conformations of unmodified and A(37)N(6)-dimethylallyl modified anticodon stem-loops of Escherichia coli tRNA(Phe)
- [153] L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, *Nat. Struct. Mol. Biol.* **2010**, *17*, 555-560. Structural aspects of messenger RNA reading frame maintenance by the ribosome
- [154] J. Gray, J. Wang, S. B. Gelvin, *J. Bacteriol.* **1992**, *174*, 1086-1098. Mutation of the miaA gene of Agrobacterium tumefaciens results in reduced vir gene expression
- [155] J. M. Durand, G. R. Bjork, A. Kuwae, M. Yoshikawa, C. Sasakawa, *J. Bacteriol.* **1997**, *179*, 5777-5782. The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of Shigella flexneri is required for expression of virulence genes
- [156] T. Suzuki, T. Suzuki, T. Wada, K. Saigo, K. Watanabe, *EMBO J.* **2002**, *21*, 6581-6589. Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases
- [157] T. Suzuki, A. Nagao, T. Suzuki, *Annu. Rev. Genet.* **2011**, *45*, 299-329. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases
- [158] H. Wang, L. Wei, C. Li, J. Zhou, Z. Li, *Oncol. Rep.* **2015**, *33*, 1089-1096. CDK5RAP1 deficiency induces cell cycle arrest and apoptosis in human breast cancer cell line by the ROS/JNK signaling pathway
- [159] R. Stupp, W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. C. Janzer, S. K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J. G. Cairncross, E. Eisenhauer, R. O. Mirimanoff, *N. Engl. J. Med.* **2005**, *352*, 987-996. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma
- [160] A. Goldbecker, A. B. Tryc, P. Raab, H. Worthmann, J. Herrmann, K. Weissenborn, *J. Neurooncol.* **2011**, *103*, 163-166. Hepatic encephalopathy after treatment with temozolomide
- [161] J. Wang, B. Yuan, C. Guerrero, R. Bahde, S. Gupta, Y. Wang, *Anal. Chem.* **2011**, *83*, 2201-2209. Quantification of Oxidative DNA Lesions in Tissues of Long-Evans Cinnamon Rats by Capillary High-Performance Liquid Chromatography–Tandem Mass Spectrometry Coupled with Stable Isotope-Dilution Method
- [162] H. Cao, Y. Wang, *J. Am. Soc. Mass Spec.* **2006**, *17*, 1335-1341. Collisionally Activated Dissociation of Protonated 2'-Deoxycytidine, 2'-Deoxyuridine, and their Oxidatively Damaged Derivatives
- [163] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**. Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine
- [164] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S.

- Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555-557. 5-Formylcytosine can be a stable DNA modification in mammals
- [165] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA
- [166] V. Busskamp, N. E. Lewis, P. Guye, A. H. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, G. M. Church, *Mol. Sys. Biol.* **2014**, *10*. Rapid neurogenesis through transcriptional activation in human stem cells
- [167] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, *145*, 423-434. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain
- [168] D. M. Franchini, E. Incorvaia, G. Rangam, H. A. Coker, S. K. Petersen-Mahrt, *PLoS One* **2013**, *8*, e82097. Simultaneous in vitro characterisation of DNA deaminase function and associated DNA repair pathways
- [169] G. Rangam, K. M. Schmitz, A. J. Cobb, S. K. Petersen-Mahrt, *PLoS One* **2012**, *7*, e43279. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud
- [170] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation
- [171] K. M. Foshay, T. J. Looney, S. Chari, F. F. Mao, J. H. Lee, L. Zhang, C. J. Fernandes, S. W. Baker, K. L. Clift, J. Gaetz, C. G. Di, A. P. Xiang, B. T. Lahn, *Mol. Cell* **2012**, *46*, 159-170. Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming
- [172] O. Habib, G. Habib, J. T. Do, S. H. Moon, H. M. Chung, *Stem Cells Dev.* **2014**, *23*, 209-218. Activation-induced deaminase-coupled DNA demethylation is not crucial for the generation of induced pluripotent stem cells
- [173] A. L. Jacobs, P. Schar, *Chromosoma* **2012**, *121*, 1-20. DNA glycosylases: in DNA repair and beyond
- [174] D. Mu, D. S. Hsu, A. Sancar, *J. Biol. Chem.* **1996**, *271*, 8285-8294. Reaction mechanism of human DNA repair excision nuclease
- [175] J. H. Hoeijmakers, *Nature* **2001**, *411*, 366-374. Genome maintenance mechanisms for preventing cancer
- [176] P. C. Hanawalt, G. Spivak, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 958-970. Transcription-coupled DNA repair: two decades of progress and surprises
- [177] I. Mellon, G. Spivak, P. C. Hanawalt, *Cell* **1987**, *51*, 241-249. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene
- [178] K. S. Sweder, P. C. Hanawalt, *Proc. Natl. Acad. Sci. USA.* **1992**, *89*, 10696-10700. Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription
- [179] J. Jiricny, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012633. Postreplicative mismatch repair
- [180] A. B. Buermeyer, S. M. Deschenes, S. M. Baker, R. M. Liskay, *Annu. Rev. Genet.* **1999**, *33*, 533-564. Mammalian DNA mismatch repair

- [181] G. M. Li, *Cell Res.* **2008**, *18*, 85-98. Mechanisms and functions of DNA mismatch repair
- [182] A. Chaudhury, P. Chander, P. H. Howe, *Rna* **2010**, *16*, 1449-1462. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles
- [183] G. Dreyfuss, V. N. Kim, N. Kataoka, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 195-205. Messenger-RNA-binding proteins and the messages they carry
- [184] W. M. Michael, P. S. Eder, G. Dreyfuss, *EMBO J.* **1997**, *16*, 3587-3598. The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein
- [185] A. Ostareck-Lederer, D. H. Ostareck, C. Cans, G. Neubauer, K. Bomsztyk, G. Superti-Furga, M. W. Hentze, *Mol. Biol. Cell* **2002**, *22*, 4535-4543. c-Src-mediated phosphorylation of hnRNP K drives translational activation of specifically silenced mRNAs
- [186] M. Chen, J. Zhang, J. L. Manley, *Cancer Res.* **2010**, *70*, 8977-8980. Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA
- [187] W. Cao, A. Razanau, D. Feng, V. G. Lobo, J. Xie, *Nucleic Acids Res.* **2012**, *40*, 8059-8071. Control of alternative splicing by forskolin through hnRNP K during neuronal differentiation
- [188] M. Shnyreva, D. S. Schullery, H. Suzuki, Y. Higaki, K. Bomsztyk, *J. Biol. Chem.* **2000**, *275*, 15498-15503. Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein
- [189] P. T. Lee, P. C. Liao, W. C. Chang, J. T. Tseng, *Mol. Biol. Cell* **2007**, *18*, 5004-5013. Epidermal growth factor increases the interaction between nucleolin and heterogeneous nuclear ribonucleoprotein K/poly(C) binding protein 1 complex to regulate the gastrin mRNA turnover
- [190] D. H. Ostareck, A. Ostareck-Lederer, M. Wilm, B. J. Thiele, M. Mann, M. W. Hentze, *Cell* **1997**, *89*, 597-606. mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end
- [191] F. Katzen, *Expert Opin. Drug Discov.* **2007**, *2*, 571-589. Gateway((R)) recombinational cloning: a biological operating system
- [192] J. L. Hartley, G. F. Temple, M. A. Brasch, *Genome Res.* **2000**, *10*, 1788-1795. DNA cloning using in vitro site-specific recombination
- [193] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, *Nat. Methods* **2014**, *11*, 319-324. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells